

# KURKKAUS KATOON

Tutkimus aikuiskoulutustutkimukseen vuonna 2006 osallistumattomista



Kuva: Elisa Nurmi 2009

Pro gradu -tutkielma

Kirsti Pohjanpää

23.9.2012

Tilastotieteen laitos

Matemaattis-luonnontieteellinen tiedekunta

Helsingin yliopisto

Tiedekunta/Osasto Fakultet/Sektion – Faculty Matemaattis-luonnontieteellinen tiedekunta		Laitos/Institution – Department Matematiikan ja tilastotieteen laitos	
Tekijä/Författare – Author Kirsti Pohjanpää			
Työn nimi / Arbetets titel – Title Kurkkaus katoon. Tutkimus aikuiskoulutustutkimukseen vuonna 2006 osallistumattomista.			
Oppiaine / Läroämne – Subject Tilastotiede			
Työn laji/Arbetets art – Level Pro gradu -tutkielma	Aika/Datum – Month and year Syyskuu 2012	Sivumäärä/ Sidoantal – Number of pages 87 sivua + 5 liitettä (15 sivua)	
Tiivistelmä/Referat – Abstract <p>Merkittävä osa länsimaisesta yhteiskunnallisesta tutkimuksesta ja tilastoinnista perustuu survey-tutkimuksiin. Jokainen otokseen valituksi tullut ei kuitenkaan syystä tai toisesta anna vastauksia tutkimuskysymyksiin. Vastauskato heikentää tutkimusaineiston laatua ja kasvattaa saatujen estimaattien virhettä.</p> <p>Tutkimus tehtiin, koska oltiin huolestuneita survey-tutkimusten kato-osuuksien kasvamisesta ja kadon vaikutuksista tutkimusten laatuun. Tutkimuksella haluttiin selvittää (a) keitä ovat ne, joita emme tutkimuksilla saa kiinni, ja (b) vastaavatko tutkimuksiin osallistumattomat samalla tapaa kuin osallistuneet.</p> <p>Selvitys tehtiin lähettämällä osalle aikuiskoulutustutkimuksesta vuonna 2006 poisjääneille uusi kontaktipyyntö tasan vuoden kuluttua. Näin syntyi AKU-laaturutkimus.</p> <p>Tiedot kerättiin käynti- ja puhelinhaastatteluin keväällä 2007. Myös laaturutkimuksesta kieltäytyneille esitettiin (jos mahdollista) muutama kysymys koskien tutkimuksiin osallistumista (ns. katohaastattelu).</p> <p>Tutkimuksen keskeisin tulos on, että jo kertaalleen tutkimuksesta suorasti tai epäsuorasti kieltäytyneitä voidaan lähestyä uudestaan hyvinkin onnistuneesti: 59 prosenttia suostui yhteistyöhön uuden kontaktin myötä. Lisäksi osoittautui, etteivät vastaamattomat juurikaan poikenneet vastanneista taustansa ja aikuiskoulutukseen osallistumisen suhteen. Laaturutkimuksen tulokset eivät tue ajatusta katoon jääneiden huonommasta yhteiskunnallisesta asemasta. Lisäksi laaturutkimukseen mukaan saadut ja siis vuotta aikaisemmin katoon kuuluneet suhtautuivat survey-tutkimukseen positiivisesti.</p> <p>Tutkimus kuuluu survey-tutkimuskenttään, ja se tuo lisää tietoa survey-tutkimuksissa pimentoon jäävästä joukosta: tutkimuskatoon kuuluvista.</p>			
Avainsanat – Nyckelord – Keywords Survey-tutkimus, vastauskato, tutkimukseen osallistuminen, logit-mallit			
Säilytyspaikka – Förvaringställe – Where deposited Matematiikan ja tilastotieteen laitos			

## Sisällys

<b>1 Johdanto: vastauskato kasvaa, tarvitaan tietoa.....</b>	<b>2</b>
<b>2 Teoriaa survey-tutkimuksen takana.....</b>	<b>4</b>
2.1 Mikä on survey? Miksi otostutkimus? .....	4
2.2 Vastauskato on ongelma.....	7
2.3 Miksi ihmiset osallistuvat tutkimuksiin?.....	8
2.4 Surveyn ominaisuudet ja osallistumispäätös .....	12
2.4.1 Tutkimuksen suunnittelu .....	13
2.4.2 Tiedonkeruumenetelmä .....	16
2.4.3 Mixed mode - eli monimenetelmällisyys .....	17
2.5 Vastauskato vaikuttaa estimaattorin harhaan ja varianssiin .....	19
2.6 Vastauskadon kasvaminen survey-tutkimuksien suuri haaste.....	20
<b>3. Tutkimusaineisto ja kuinka siihen päädyttiin .....</b>	<b>22</b>
3.1 Kurkkaus katoon: päätettiin tehdä AKU-laaturutkimus.....	22
3.3 Laaturutkimuksen otos oli vastauskatoa vuonna 2006 .....	23
3.3 Tiedonkeruun kulku .....	25
3.4 Tiedonkeruu ja haastattelijatyö.....	27
3.5 Käytetyt menetelmät .....	28
3.6 Johdatus tuloksiin .....	39
<b>4. Tulokset: mitä kato kertoo meille? .....</b>	<b>41</b>
A Tutkimukseen osallistuneet .....	41
4.1 Valmiutta osallistua on: vastausosuus 59 prosenttia .....	41
4.2 Kato ei sittenkään huono-osaisempaa.....	42
4.3 Jakaumat muistuttavat toisiaan .....	44
4.4 Laaturutkimukseen osallistuneilla parempi työmarkkina-asema .....	48
4.5 Haastatteluun osallistuneet äänestivät eduskuntavaaleissa .....	50
4.6 Tutkimukseen osallistuminen logit-mallituksen valossa.....	51
B Katokysymykset .....	56
4.7 Tilastokeskus tunnetaan ja arvostusta löytyy.....	56
4.8 Tutkimuksia pidetään tärkeinä.....	57
4.9 Päätös osallistua syntyi kirjeen ja haastattelijan soiton myötä.....	60
4.10 Aikaisempi yhteydenotto muistettiin.....	62
4.11 Pitäisikö haastatteluun osallistumisesta palkita?.....	63
C Aikuiskoulutustutkimuksen keskeiset indikaattorit .....	65
4.12 Aikuiskoulutukseen osallistutaan samassa määrin .....	65
4.13 ...mutta aikuiskoulutustarpeita koetaan enemmän.....	67
4.14 Osallistuminen aikuiskoulutukseen logit-mallituksen mukaan.....	68
<b>5. Yhteenveto ja pohdinta .....</b>	<b>70</b>
5.1 Vastaus tutkimuskysymyksiin.....	70
5.2 Tulosten suhde teoriaan.....	74
5.3 Riippuuko saalis haavista?.....	78
Lähteet .....	82
Liite 1. Katokysymykset AKU-laaturutkimuksessa .....	ii
Liite 2. Luettelo aikuiskoulutusmuodoista (lähetetty saatekirjeen mukana) .....	v
Liite 3. AKU 2006:n katokoodit ja AKU-laaturutkimuksen otos .....	vi
Liite 4. AKU-laaturutkimuksen saatekirjeet: ei-tavoitetut ja kieltäytyneet .....	vii
Liite 5. Logit-mallitus: muuttujat ja mallit .....	ix

## *1 Johdanto: vastauskato kasvaa, tarvitaan tietoa*

Merkittävä osa länsimaisesta yhteiskunta-, sosiaali- ja lääketieteellisestä tutkimuksesta ja tilastoinnista perustuu survey-tutkimuksiin. Näin siksi, että otostutkimus on tilastotieteellisesti hyvin perusteltu ja kansantaloudellisestikin tehokas tapa saada tietoa ilmiöistä, jotka eivät rekisteritiedoilla ole saavutettavissa. Kun kotitalouksien kulutusta, kansalaisten työtuntien määrää, ajankäytön jakautumista ja aikuiskoulutukseen osallistumista ei hallinnollista aineistoista löydy, on kysyttävä ihmisiltä itseltään. Kaikilta kansalaisilta ei tarvitse kysyä, voimme ottaa heistä edustavan otoksen ja laskea estimaatit tätä kautta. Jokainen otokseen valituksi tullut ei kuitenkaan syystä tai toisesta anna vastauksia tutkimuskysymyksiin. Tämä vastauskato heikentää tutkimusaineiston laatua ja kasvattaa saatujen estimaattien virhettä.

Tämä tutkimus tehtiin, koska oltiin huolestuneita survey-tutkimusten kato-osuuksien kasvamisesta ja vastauskadon vaikutuksista tutkimusten ja tilastoinnin laatuun. Ilmiö on yleiseurooppalainen ja jatkunut koko 2000-luvun ajan.

Huoli konkretisoitui Tilastokeskuksen aikuiskoulutustutkimuksen osalta vuonna 2006. Tuolloin tutkimuksen vastausosuus oli 65 prosenttia, sinänsä siedettävä, mutta runsaat 9 prosenttiyksikköä korkeampi kuin kuusi vuotta aikaisemmin toteutetussa tutkimuksessa. Vastauskadon kasvun seurauksena päätettiin Tilastokeskuksessa selvittää, keitä meiltä jää katoon. Perinteisesti hyvään katoanalyysiin kuuluu kadon rakenteen tarkastelu käytettävissä olevien, lähinnä rekisteritietojen avulla (ikä, sukupuoli, asuinpaikka, koulutus ja niin edelleen). Tällä kertaa haluttiin tietoa myös vastaamattomaksi jääneiden muista ominaisuuksista. Keskeinen tutkimushypoteesi oli, keitä he ovat ja osallistuvatko he aikuiskoulutukseen samassa määrin kuin aikuiskoulutustutkimukseen suostuneet.

Selvitys tehtiin lähettämällä osalle aikuiskoulutustutkimuksesta vuonna 2006 poisjääneille uusi kontaktipyyntö tasan vuoden kuluttua. Näin syntyi aikuiskoulutustutkimuksen 2006 laatututkimus, jonka toteutustapaa ja tuloksia käsitellään tässä tutkimusraportissa. Tutkija on suunnitellut tutkimusasetelman, muotoillut ja koodannut tiedonkeruulomakkeet, valmistanut tiedonkeruumateriaalin,

kouluttanut haastattelijat, tarkistanut ja editoinut saadun tutkimusaineiston sekä analysoinut tulokset. Tutkimuksen tuloksia on raportoitu Tilastokeskuksen sisäisessä raportissa (Pohjanpää 2009a), Hyvinvointikatsauksen artikkelissa (Pohjanpää 2010) sekä tieteellisessä konferenssissa (Pohjanpää 2011). Päätös käyttää tutkimusta tähän pro gradu -tutkielmaan syntyi myöhemmin.

Tutkimus kuuluu survey-tutkimuskenttään, ja se tuo lisää tietoa survey-tutkimuksissa pimentoon jäävästä joukosta: tutkimuskatoon kuuluvista. Kurkataanpa katoon.

## 2 Teoriaa survey-tutkimuksen takana

### 2.1 Mikä on survey? Miksi otostutkimus?

Survey on yleisnimi menetelmille, joilla hankitaan yksilökohtaista kvantitatiivista tietoa väestöstä ja sen yksiköistä (esimerkiksi yksilö, kotitalous tai yritys). Useimmiten survey perustuu otostutkimukseen, mutta se voi perustua myös kokonaistutkimuksiin.

Surveyssä voidaan käyttää erilaisia tiedonkeruumenetelmiä, kuten puhelin- tai käyntihaastattelua (yleisimmin käytetään termejä *CATI Computer Assited Telephone Interviewing* ja *CAPI Computer Assited Personal Interviewing*) tai lomakekyselyä. Nykyisin web-tiedonkeruu on syrjäyttämässä perinteisen postikyselyn. Tutkimusaineiston laadun ja kadon kannalta on oleellista, täyttääkö lomakkeen *haastattelija* vai *vastaaaja itse*. Tiedonkeruussa (yksittäinen haastattelu, tiedonkeruuprosessi) voidaan käyttää kahta tai useampaa eri tiedonkeruumenetelmää. Tällöin kyseessä on niin sanottu mixed mode -tiedonkeruu, jonka suomenkielisenä terminä käytetään usein *monimenetelmällisyys*. Tiedonkeruumenetelmän valintaan vaikuttavat tutkimusaihe, kohdejoukko, kustannukset, eri tiedonkeruumenetelmiin liittyvät mahdollisuudet (resurssit, osaaminen) sekä tutkimusperinne ja tottumus. (Katso myös luvut 2.4.2 ja 2.4.3.) Termejä tiedustelu ja syväluotaus (Laaksonen 2010) on ehdotettu surveyn suomennokeksi, mutta kummankaan käyttö ei ole vakiintunut suomalaiseen survey-tutkimuskenttään.

Survey koskettaa harvoin koko perusjoukkoa. Useinkaan kaikkia perusjoukon tapauksia (esimerkiksi koko väestöä) ei ole mahdollista eikä edes tarpeen tutkia. Käytännöllisempää on valita perusjoukosta edustava osajoukko, otos, johon tutkimus suunnataan. Otos voidaan poimia joko yksinkertaisena satunnaisotantana (*SRS, simple random sampling*) tai systemaattista (*SYS, systematic sampling*), ositettua (*STR, stratified sampling*) tai ryväotantaa (*CLU, cluster sampling*) käyttäen tai otanta voidaan tehdä otosyksikön koon mukaan (*PPS, probabilities proportional to size*). Satunnaisotannassa jokaisella perusjoukon yksilöllä on sama todennäköisyys tulla valituksi otokseen, kun taas muissa edellä mainituissa otannoissa voidaan vaikuttaa yksilön valikoitumistodennäköisyyteen ja otoskehikon muodostamiseen.

Henkilösurveyssä käytetään yleisesti otantakehyksenä väestörekisteristä poimittua satunnaisotosta edustamaan Suomessa asuvia tietynikäisiä henkilöitä. (Pahkinen & Lehtonen 1989.)

Otannassa tutkittavien yksiköiden todennäköisyys tulla valituksi voidaan etukäteen laskea ja sen perusteella lasketaan myöhemmin havaintoaineiston tilastolliset *estimaatit*, jotka kuvaavat tutkittavan ilmiön yleisyyttä perusjoukossa. Otostutkimuksessa tutkimusnäyte eli otos muodostaa pienoismallin väestöstä ja estimaatit vastaavat tutkittavan ilmiön arvoja.

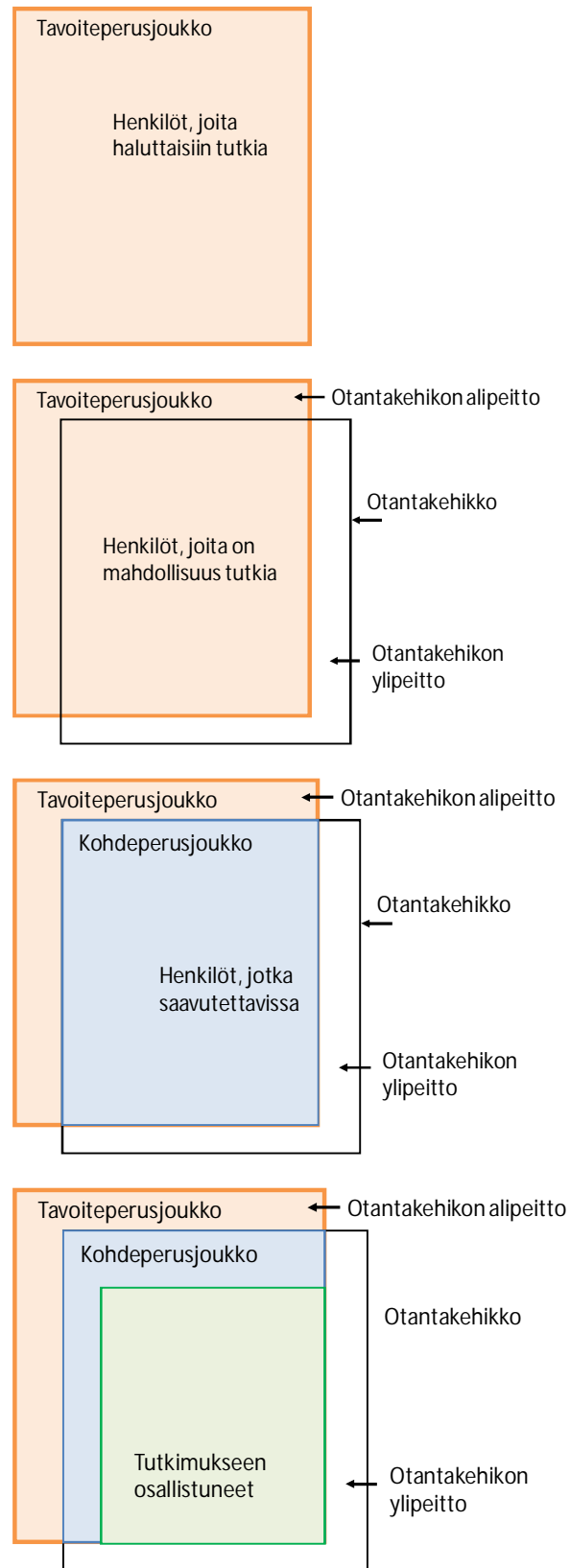
Otosasetelma on olennainen osa tutkimusongelman määrittelyä. Hyvin suunnitellulla otannalla voidaan hallita monimutkaisia ilmiöitä. Otannan käyttö edellyttää, että tutkijalla on olemassa tietoa perusjoukosta ja että perusjoukon kehikko tai luettelo, josta otanta voidaan suorittaa, on saatavilla. Suomessa ja muissa Pohjoismaissa on hyvät mahdollisuudet tehdä otostutkimuksia, mutta tilanne ei ole yhtä hyvä muualla Euroopassa ja Amerikassa.

Otanta kuvaa teoreettista tilannetta, mutta käytännössä hyvissäkin rekistereissä on puutteita, päivityksen viivettä sekä muita häiritseviä tekijöitä, jotka heikentävät teorian ja konkretian kohtaamista. *Tavoiteperusjoukoksi* (*target population*) kutsutaan sitä väestön osaa, josta tutkimuksella halutaan tietoja. Tavoiteperusjoukkona voi olla esimerkiksi 16–64-vuotiaat Suomessa asuvat<sup>1</sup>. *Otoskehikko* rajaa tietyn osa tavoiteperusjoukosta ulos (ilmiötä kutsutaan *alipeitoksi*, *undercoverage*), mutta samalla se voi sisältää ylimääräisiä henkilöitä (*ylipeitto*, *overcoverage*). Yli- ja alipeiton huomioimisen myötä saamme *kohde- tai kehikkoperusjoukon* (*frame population*), joka voi olla esimerkiksi 16–64-vuotias väestö pois lukien laitospöytä. (Kuvio 2.1.)

---

<sup>1</sup> Tässä teoreettinen käsite ”suomalaiset” on jo konkretisoitunut mitattavissa olevaan käsitteeseen ”Suomessa asuvat”.

**Kuvio 2.1** Tavoite- ja kohdeperusjoukko, otantakehikko ja tutkimusaineisto survey-tutkimuksessa.





Otostutkimuksessa pyritään estimoimaan tiettyjä tutkittavan perusjoukon ominaisuuksia esimerkiksi työttömyyttä tai aikuiskoulutukseen osallistumista (parametrina aikuiskoulutuksen osallistumisaste). Parametrien arvioina eli estimaatteina käytetään tutkimusaineistosta laskettuja tunnuslukuja (esimerkiksi aikuiskoulutukseen osallistuneiden keskimääräinen osuus otoksessa). Tunnuslukujen laskentakaavoja kutsutaan estimaattoreiksi.

Ensimmäinen nykymuotoinen otostutkimus on Norjassa 1890-luvulla tehty haastattelututkimus, joka koski norjalaisten sosiaalisia oloja (Kuusela 2009). Tilastokeskuksen kulutustutkimuksen alkuna pidetään Veera Hjeltin haastatteluja 1900-luvun alussa ja työvoimatutkimus perustuu 1959 tehtyyn laskentaan (Keinänen 2009). Nykyisin survey on laajasti käytössä ja se on perustana usealle Suomen viralliselle tilastolle, sitä käytetään tutkimuslaitoksissa, korkeakouluissa ja yliopistoissa tieteellisen tutkimuksen aineistonkeruumenetelmänä.

## 2.2 Vastauskato on ongelma

Vaikka survey on menetelmänä erinomainen ja laajasti hyödynnetty, liittyy siihenkin tiettyjä puutteita. Yksi keskeinen survey-tutkimuksen heikentäjä on tutkimuskato: vain harvoin kaikki otokseen valitut yksiköt osallistuvat tutkimukseen. Tutkimuksesta poisjäämistä mitataan kato- (*nonresponse rate*) ja vastausosuuksilla (*response rate*). Vastausosuus on yksinkertaisesti vastanneiden lukumäärä jaettuna otoksen koolla (brutto, netto). Vastauskato voi kohdistua koko tutkimukseen, jolloin puhutaan yksikkökadosta (*unit nonresponse*). Jos vastaaja jättää vastaamatta yksittäisiin kysymyksiin, on kyse eräkadosta (*item nonresponse*)<sup>2</sup>.

Tutkimuskato lisää aineistosta laskettujen estimaattorien harhaa sekä pienentää aineiston kokoa ja edustavuutta. Tilastojen ja tutkimusaineistojen tulee olla edustavia myös väestöryhmittäisten analyysien osalta. Aineiston pieneneminen heikentää analysointimahdollisuuksia, nostaa tiedonkeruun yksikkökustannuksia sekä hankaloittaa tiedonkeruuprosessin suunnittelua. Esimerkiksi tarvittavien haastattelijoiden määrä ja tutkimuksen kentällä oloaika riippuvat siitä, kuinka

---

<sup>2</sup> Tässä tutkielmassa ei pohdita eräkadon aiheuttamaa tutkimusaineiston laadun heikkenemistä, koska tarkastelun kohteena olleiden tutkimusten (aikuiskoulutustutkimus ja AKU-laaturitkimus) osalta eräkatoa ei juuri ilmennyt.

helposti otoshenkilöt saadaan tavoitettua ja houkuteltua mukaan tutkimukseen. (Biemer & Lyberg 2003, 95).

Survey-tutkimuksen laatuun vaikuttavat kadon lisäksi myös monet muut seikat: otoksen edustavuus, tietojen tallentamisen virheettömyys, haastattelujen onnistuminen, tarkistuksien ja korjausten onnistuminen ja monet jo lomakkeen suunnittelun vaiheet (tutkimusongelman muotoileminen haastattelukysymyksiksi, termien ja käsitteiden yksiselitteisyys, lomakkeen koodauksen virheettömyys ja niin edelleen). Edellä mainitut seikat otoksen edustavuutta lukuun ottamatta rajataan tämän tutkimuksen ulkopuolelle.

Suuri vastauskato ei sinällään ole automaattisesti huonon laadun merkki eikä suuri vastausosuus hyvä laadun tae (Groves & Couper 1998; Groves 1998; Groves 2006). Kadon vinous eli se, kuinka tasaisesti tai epätasaisesti tutkimuksesta poisjääminen on jakautunut perusjoukon yksiköiden kesken, on keskeinen tutkimusaineiston laadun kannalta. Epätasaisesti jakautunut vastauskato heikentää aineiston laatua ja sitä mitataan katovirheellä (*nonresponse bias*). Yksinkertaistaen voidaan todeta, että 50 prosentin vastausosuus on melko huono, mutta aivan mahdoton se on, jos väestötutkimuksessa vain naiset vastaavat ja miehet eivät. Jos vastauskato on 50 prosenttia sekä naisten että miesten, nuorten ja iäkkäiden, maaseudulla ja kaupungissa asuvien ja eri koulutustaustan omaavien keskuudessa, niin sen vaikutus tutkimusaineiston laatuun ei ole enää yhtä huolestuttava.

### 2.3 Miksi ihmiset osallistuvat tutkimuksiin?

Kun vastaajaehdokas (otoshenkilö) tekee päätöstään tutkimukseen osallistumisesta, on kyse samoista sosiaalista ja psykologista lainalaisuuksista kuin ihmiselämässä yleensäkin. Tutkimukseen valittu on tavalla tai toisella vuorovaikutuksessa tutkimuksen tekijän kanssa.

Markkinoinnin ja psykologian professorin Robert Cialdinin käsitystä kuudesta ”vaikuttamisaseesta” (*weapons of influence*) on hyödynnetty myös yritettäessä ymmärtää, kuinka ihmiset tekevät päätöksensä osallistua tai olla osallistumatta tieteelliseen tutkimukseen (Groves & Couper 1998, 32–34; katso myös Biemer & Lyberg 2003, 96–101; Minska bortfallen 1998).

a) **Vastavuoroisuus** (*reciprocation*) perustuu yleiseen normiin siitä, että positiiviseen ja miellyttävään käytökseen vastataan positiivisesti, ystävällisyys palkitaan ystävällisyydellä. Haastattelijoiden rooli tutkimuslaitoksen edustaja on tässä ilmeinen.

b) **Johdonmukaisuus** (*consistency*) ilmenee ihmisten haluna olla mielipiteidensä, uskomustensa, arvojensa ja tekojensa suhteen johdonmukaisia. Arvojen ja asenteiden muuttuminen on yleensä hidasta.

c) **Sosiaalinen vahvistus** (*social validation*) antaa oikeuden omille arvoille ja toiminnalle. Sosiaalisen vahvistuksen hakemiseen perustuu samaistuminen erilaisiin viiteryhmiin, ja tätä hyödynnetään erityisesti tietyille väestöryhmille kohdennetuissa survey-tutkimuksissa ja saatekirjeiden muotoilussa.

d) **Auktoriteetti** (*authority*) liittyy tahoihin, jotka edustavat henkilölle erityisasemaa ja arvovaltaa ja saavat siten erityiskohtelua. Osallistumispäätös voi perustua siihen, että tutkimuksen toteuttaa tai sitä suosittaa vastaajan arvostama instituutio, laitos tai tutkija.

e) **Ainutkertaisuus** (*scarcity*) kiehtoo ihmisiä, ja ainutkertaisiin tilaisuuksiin tartutaan usein tomerasti.

f) **Mieltymys** (*liking*) auttaa vuorovaikutuksen ja yhteistyön synnyssä. Survey-tutkimuksissa onkin ensikontaktilla (saatekirje, haastattelijan yhteydenotto) suuri merkitys.

Osallistumispäätöstä voidaan tarkastella myös **sosiaalisen vaihdon teorian** (*social exchange theory*) ja siitä johdetun **rationaalisen valinnan teorian** (*rational choice theory*) kautta. Sosiaalisen vaihdon teoria on yksi keskeisimmistä ihmisten yhteistoimintaa käsittelevistä teorioista sosiaalipsykologiassa<sup>3</sup>. Teoria perustuu käsitykseen, että ihmiset tekevät päätöksiä kustannus–hyöty -näkökulmasta. Ihmiset vertaavat (kuvitteellista) saamaansa hyötyä siihen, kuinka paljon asia tai ilmiö aiheuttaa heille kustannuksia. (Katso esim. Dillman et al. 2002, 8.)

---

<sup>3</sup> Sosiaalisen vaihdon teorian isänä pidetään George Homansia (*Social Behavior: Its Elementary Forms* 1961).

Kustannuksia survey-ympäristössä aiheuttavat esimerkiksi vaikea tai epäkiinnostava tutkimusaihe, vastaajan (todelliset tai arvioidut) kyvyt selvittää vastaamisprosessista sekä vastaajan motivoituneisuus. Kustannusten ylittäessä hyödyn seuraa tutkimuksesta kieltäytyminen. Ilmiötä esiintyy myös kysymystasolla – vaikeissa kysymyksissä ilmenee enemmän osittaiskatoa. (Esim. Alwin 1991.)

Osallistumispäätös voi myös perustua *haluun auttaa*. Tähän vedotaan usein jo saatekirjeessä. Tutkimukset osoittavat, että pelkän lauseen ”Osallistumisenne todella auttaisi meitä” lisääminen saatekirjeeseen nostaa vastausaktiivisuutta (Mowen & Cialdini 1980).

Auttamishalun heräämiseen vaikuttaa oleellisesti edellä mainitut kuusi psykologista elementtiä. Se on yleensä aina vastavuoroista, sen takana on joukko arvoja ja asenteita ja tämän takia ihminen pyrkii myös auttamisen suhteen toimimaan johdonmukaisesti, kaipaa sosiaalista vahvistusta sekä pyrkii yhdenmukaisuuteen. Ihmisiä voidaan houkutella auttamaan vetoamalla auktoriteettiin tai siihen, että juuri sinun apuasi tarvitaan.

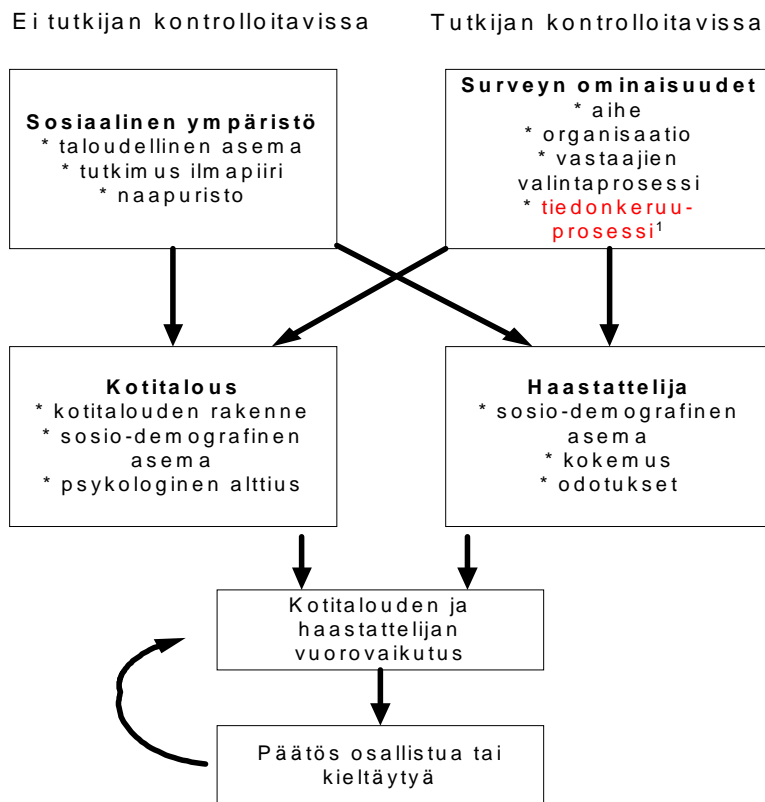
Ihmisten kyky ja halukkuus – tai oikeastaan *haluttomuus muuttaa mielipiteitään* on myös tärkeä psykologinen ominaisuus, joka vaikuttaa tutkimukseen osallistumisprosessissa. Mielipiteen muuttamisella on aina hintansa ja muuttamishalukkuus riippuu oleellisesti siitä, kuinka vahva alkuperäinen mielipide oli, kuinka vahvoja argumentteja muuttamisen puolesta tarjotaan ja millä tapaa ne esitetään. Myös harkinta-ajalla on merkitystä. Joutuessaan tilanteeseen, jossa omien arvojen ja mielipiteiden tarkistaminen ja mahdollinen muuttaminen voisi olla ajankohtaista, kohtaa ihminen hämmennystä. Hämmennyksen voimakkuus ja ihmisen kyky käsitellä sitä vaikuttavat lopputulokseen. Tämän takia hyvin perustellut argumentit (jotka mahdollisesti sivuavat jollakin kohtaa vastaajan omaa arvomaailmaa) menevät helpommin läpi ja tuovat todennäköisimmin halutun lopputuloksen.

Osallistumispäätökseen vaikuttavat monet seikat, muun muassa tutkimuksen aihe, tutkimusorganisaatio, tutkimuksen ajankohta ja perusjoukko, haastattelijan toiminta ja habitus, haastateltavan henkilökohtaiset ominaisuudet, osaaminen, asenteet ja arvot

sekä hänen aikaisemmat kokemukset vastaavista tilanteista (Groves, Cialdini & Couper 1992). Myös sattuma saattaa vaikuttaa osallistumispäätökseen ajankohdan ja vastaajan hetkellisen mielialan kautta. Tutkija pystyy vaikuttamaan vain murto-osaan osallistumispäätökseen vaikuttavista seikoista (Groves ja Couper 1998, kuvio 2).

Alkuperäisesti kuvioista puuttui yksi keskeinen surveyn ominaisuus, johon tutkija voi vaikuttaa ja joka vaikuttaa otoshenkilön osallistumispäätökseen. Kutsun sitä *tiedonkeruuprosessiksi* tai *tiedonkeruun käytännöiksi* (*survey protocol*) ja se käsittää muun muassa tutkimuksen ajankohdan, yhteydenottotavan (saatekirje, haastattelijan kontaktit) sekä niiden määrän, osallistumispalkkion ja tiedonkeruumenetelmän. Nämä kaikki ovat seikkoja, jotka suunnitellaan hyvin tutkimusta valmistettaessa. Suunnittelussa huomioidaan tutkimuksen tavoitteiden lisäksi se, kuinka kustannustehokkaasti saataisiin mahdollisimman moni osallistumaan tutkimukseen. Tiedonkeruuprosessin suunnittelun yhtenä keskeisenä tavoitteena on myös vaikuttaa otoshenkilöiden päätösprosessiin.

**Kuvio 2.2** Tutkimukseen osallistumisen päätösprosessi



Lähde: Groves & Couper 1998, 39.

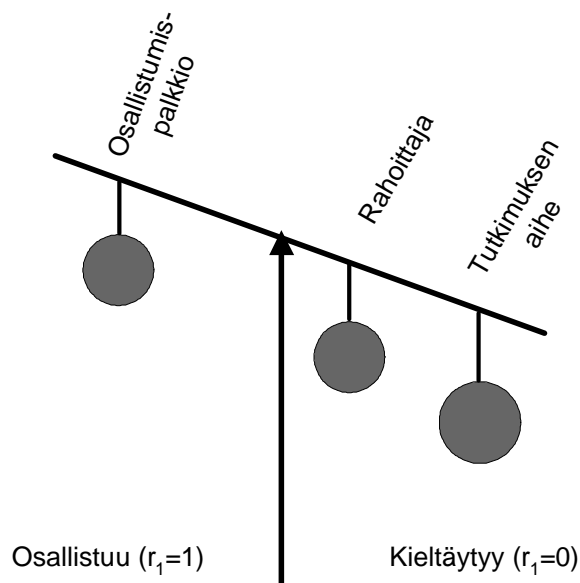
<sup>1</sup> Tiedonkeruuprosessi on tutkijan oma lisäys tähän kuvioon.

## 2.4 Surveyn ominaisuudet ja osallistumispäätös

Surveyn ominaisuudet vaikuttavat kansalaisten päätökseen osallistua tutkimuksiin. Groves, Singer ja Corning kehittivät niin sanotun *Leverage–Salience* -teorian<sup>4</sup> kuvaamaan näitä vaikutuksia (Groves et al. 2000; Groves et al. 2004). Teoria perustuu ajatukseen, että tutkimuksen ominaisuuksilla (kuten esimerkiksi aiheella, saatekirjeillä ja palkkioilla) voilla olla erilaisia vipuvaikutuksia (*leverage*) eri henkilöille (kuvio 2.3). Surveyn ominaispiirteet vetävät henkilön päätöstä osallistumisen tai tutkimuksesta kieltäytymisen suuntaan. Tähän vaikuttaa se, kuinka suuri merkitys (*salient*) näillä surveyn ominaisuuksilla on yksittäisille henkilöille. Vaikutus voi riippua henkilön iästä, sukupuolesta, koulutuksesta ja jopa vastaamishetken mielialasta tai elämäntilanteesta.

Tutkijat korostavat, että käyttämällä erilaisia surveyn ominaisuuksia voidaan eri vastaajaryhmiä helpottaa osallistumaan tutkimukseen. Monet survey-tutkimukset 2000-luvulla ovat vahvistaneet teoriaa (Groves et al. 2006; Adua & Sharp 2010).

**Kuvio 2.3** Leverage–Salience -teorian periaate



Lähde: Groves et al. 2000

<sup>4</sup> En ole törmännyt Leverage–Salience -teorian suomennokeeseen.

### 2.4.1 Tutkimuksen suunnittelu

Seuraavaksi käsitellään joitakin sellaisia surevyn ominaisuuksia, joilla pyritään helpottamaan kansalaisten päätöstä osallistua survey-tutkimuksiin. Osa keinoista on vain pieniä valintoja tutkimusprosessia suunniteltaessa, mutta osa keinoista vaatii lisää resursseja tutkimukseen. On selvää, että monessa tapauksessa vastausosuuden kasvattaminen kasvattaa tiedonkeruukustannuksia (Lynn et al. 2002; Teitler et al. 2003).

*Tutkimuksen aihe* määräytyy tieteellisin perustein ja tutkimusyhteisön intressien mukaan. Tutkimusaine vaikuttaa tutkimukseen osallistumispäätökseen (Adua & Sharp 2010). Tutkijan tehtävänä on tarjoilla tutkimusaihe vastaajakokelaille siten, että kiinnostus syntyy. Usein tutkimuksen aihe on kiinteästi yhteydessä myös *tutkimuksen toteuttaja-* ja/tai rahoittajatahoon. Näiden tahojen arvovaltaisuus ja maine vaikuttavat kansalaisten päätökseen osallistua tutkimukseen (Groves & Couper 1998, 275).

Tutkimuksen *pituus*, joka vastaajakokelaille ilmoitetaan usein haastattelun kestona, on yksi keskeinen seikka, joka vaikuttaa henkilöiden innostukseen osallistua. Tiedonkeruulomakkeen laajuus onkin usein kompromissi tutkijan mielenkiinnon ja vastausrasitteen (oletettu, koettu ja todellinen) välillä.

Tutkimuksen *ajankohta* määräytyy usein tutkimuksen aiheen ja tutkimuksessa tarvittavien resurssien käytettävissä oloajan mukaan. Luonnollisesti myös vastaajien ajankäyttö huomioidaan tutkimusta suunniteltaessa niin, että esimerkiksi tiedonkeruu ei ajoitu loma- eikä keskeisten juhlapyhien aikoihin. Tutkimukset eivät kuitenkaan tue ajatusta, että kesä ja joulukuu olisivat automaattisesti huonompia tiedonkeruukuukausia (Losch et al. 2002).

*Saatekirjeiden* vaikutus vastaushalukkuuteen on kiistaton (De Leeuw et al. 2007), sillä monet tutkimukset ovat osoittaneet niiden nostavan tutkimuksen vastausosuutta (esim. Link & Mokdad 2005; Mann 2005; Hembroff et al. 2005; Heerwegh 2005). Saatekirjeiden käyttö onkin vakiintunut osaksi hyvää suomalaista ja eurooppalaista

tutkimusperinnettä<sup>5</sup>. Meillä Suomessa saatekirjeiden lähettämistä pidetään yhtenä tutkimuseettisenä periaatteena. Saatekirjeissä voidaan muun muassa perustella, miksi tutkimus on tärkeä sekä tuoda julki tutkimuksen toteuttaja- ja rahoittajatahot. Myös muistutuskirjeiden muodolla, määrällä ja ajoituksella on vaikutusta survey-osallistumispäätökseen.

*Yhteydenottojen määrällä* on vaikutusta siihen, kuinka suuri osa kansasta saadaan osallistumaan survey-tutkimuksiin. Jos henkilöä yritetään tavoitella vain muutamia kertoja, on onnistumiselle annettu vain pieni mahdollisuus. Jossakin vaiheessa yhteydenottojen lisääminen ei kuitenkaan enää kasvata vastausosuutta. Lisäksi yhteydenottojen määrää on tarpeen rajoittaa tiedonkeruajan ja -kustannusten takia.

Selkeää yleispätevää sääntöä, kuinka monta kontaktia tarvitaan, ei ole mahdollista määritellä. Määrät vaihtelevat tutkimuksen aiheen, tiedonkeruajan ja -tavan, ja jopa otoshenkilöiden kansalaisuuden mukaan. Romanov ja Nir (2010) havaitsivat, että Israelissa kulutustutkimuksessa kontaktien vähentäminen arabien osalta kolmeen ja juutalaisten osalta neljään kontaktiin pudotti vastausosuutta 12 prosenttiyksikköä, mutta ei vaikuttanut saadun tutkimusaineiston keskeisten ydinmuuttujien laatuun. Vastaavanlaisia tuloksia on saatu muuallakin (Curtin et al. 2000; Keeter et al. 2000 & 2006).

*Osallistumispalkkiolla (incentive)* pyritään nostamaan otokseen valitun henkilön innostusta osallistua tutkimukseen. Palkkiona on yleisesti käytetty pientä tavara- tai rahalahjaa. Osallistumispalkkioina käytetään myös osallistuneiden kesken tehtävää arvontaa, jossa pääpalkintona voi olla esimerkiksi tv tai tietokone.

Palkkiolla on osoitettu olevan vastausosuutta kasvattava vaikutus (Groves et al. 2000; Singer et al. 2000; Trussell & Lavrakas 2004; Groves et al. 2004; Groves et al. 2006; Jäckle & Lynn 2008). Lisäksi palkitsemisella on onnistuttu saamaan tutkimukseen mukaan vaikeasti tavoitettavissa olevia ja helposti tutkimuksesta kieltäytyviä väestöryhmiä (Barón et al. 2009).

---

<sup>5</sup> Maissa, joissa ei voida tehdä rekistereihin perustuvaa otostutkimusta, ei saatekirjeiden lähettäminen ennen tutkimusta ole aina mahdollista osoite- ja henkilötietojen puutteen takia.



Palkkio tuntuu toimivan parhaiten, kun se lähetetään etukäteen saatekirjeen mukana (Dillman 2007; Trussell & Lavrakas 2004; Brennan & Charbonneau 2009; Tourangeau et al. 2010). Etukäteen annetulla palkkiolla voi olla jopa suurempi vaikutus osallistumisaktiivisuuteen kuin tutkimuksen aiheella (Adua & Sharp 2010; Tourangeau et al. 2010; Groves et al. 2004).

Palkitsemisen hyödyistä survey-tutkimuksessa on keskusteltu paljon. Vaikka palkkiolla saadaan kasvatettua vastausosuutta ja pienennettyä katovirhettä, voi palkkio kasvattaa aineiston mittausvirhettä. Näin kävi esimerkiksi äänestämistä koskevassa tutkimuksessa, jossa onnistuttiin palkkioiden avulla houkuttelemaan erityisesti ei-äänestäjiä mukaan tutkimukseen. Heillä oli kuitenkin suurempi taipumus vastata virheellisesti (koskien omaa äänestyskäyttäytymistään) kuin vaaleissa äänestäneillä. (Tourangeau et al. 2010.) Palkkion etukäteen saaneiden on osoitettu myös vastaavat tutkimuskysymyksiin muita hieman positiivisemmin (Singer et al. 2000). Tutkimukset ovat myös osoittaneet päinvastaista: palkkioiden käyttö ei heikentänyt tutkimuksen laatua (Davern et al. 2003).

Jos palkkioilla onnistutaan saamaan mukaan tutkimuksiin epäilevästi suhtautuvia ja jos he motivoituvat tarpeeksi vastataksaan oikein ja tunnollisesti, ovat palkkiot paikallaan. Jos kuitenkin käy niin, että palkkion houkuttelemat vastaavat puutteellisesti tai muutoin virheellisesti, haittaavat palkkiot tutkimusta.

On myös pohdittu, sopiiko palkkio tutkimuslaitosten ja tilastoviranomaisten tekemiin survey-tutkimuksiin. Ehtona on pidetty vähintään sitä, että palkkio on annettava kaikille otokseen valituille (Biemer & Lyberg 2003, 109). Palkkion haittana on myös se, että pienikin palkkio voi nostaa tiedonkeruukustannuksia kymmeniä-, jopa satojatuhansia euroja otoksen suuruuden kautta.

*Tiedonkeruuprosessilla* tai *tiedonkeruun käytännöillä* (*survey protocol*) tarkoitetaan kokonaisuutta, joka alkaa otoshenkilön valintamenettelyistä ja päättyy haastatteluun. Siihen kuuluvat saatekirjeet, haastattelijoiden koulutus, kontaktitavat, -määrät ja niiden ajoitus sekä tiedonkeruumenetelmä. Peytchiev kumppaneineen (2009) osoitti, että yhteydenottojen määrää muuttamalla saatiin kasvatettua vastausosuutta, mutta vasta yhdessä tiedonkeruuprosessin muutoksen kanssa onnistuttiin tutkimukseen

houkuttelemaan myös vaikeasti tavoitettavia (ja tutkimukseen erilailla vastaavia). Näin saatiin katovirhettä pienennettyä.

### 2.4.2 Tiedonkeruumenetelmä

Myös tiedonkeruumenetelmällä pyritään vaikuttamaan vastaajien mahdollisuuksiin ja innostukseen osallistua tutkimukseen. Tiedonkeruumenetelmän valintaan vaikuttavat myös monet muut syyt: tutkimusaihe, kohdejoukko, kustannukset, eri tiedonkeruumenetelmiin liittyvät mahdollisuudet (resurssit, osaaminen) sekä tutkimusperinne ja tottumus (De Leeuw 2007 & 2008).

Tiedonkeruumenetelmän valinta on yksi tutkimusprosessin keskeisimpiä ja suurimpia päätöksiä. Tiedonkeruumenetelmällä on suuria vaikutuksia tutkimuksen kustannuksiin, laatuun ja aikatauluun sekä toteuttamismahdollisuuksiin (esimerkiksi koulutettujen haastattelijoiden saanti).

On joukko seikkoja, jotka selkeästi erottavat eri tiedonkeruumuodot toisistaan. Vastaamisprosessin erot liittyvät haastattelijan läsnäoloon ja siihen, perustuuko kysymyksen ymmärtäminen ja vastaaminen visuaalisuuteen vai kuuloon (puhumiseen). Myös kysymysten muotoilu vaihtelee tiedonkeruumenetelmien välillä (itse täytettävä lomake tai haastattelijä täyttää). (Dillman et al. 2009, 310–311.)

Puhelin- ja käyntihaastatteluissa haastattelun kulkuun vaikuttaa haastattelijan läsnäolo. Hän voi auttaa vastaajaa ymmärtämään kysymyksiä, mutta toisaalta haastattelijan kuullen ei ehkä ”kehdata” olla täysin rehellisiä (negatiivinen haastattelijavaikutus, sosiaalinen suotavuus, *social desirability*).

Käyntihaastattelu eroaa puhelinhaastattelusta muun muassa siinä, että siinä on mahdollista käyttää apuna visuaalisia elementtejä (esimerkiksi vastauskortteja). Lisäksi käyntihaastattelun kesto voi olla pidempi. Myös yksittäiset kysymykset voivat käyntihaastattelusta olla hiukan monimutkaisempia ja pidempiä.

Itse täytettävissä lomakkeissa (postikysely, Internet-tiedonkeruu) ei ole haastattelijavaikutusta, mutta toisaalta vastaajan tulee itse selvittää koko vastaamisprosessista. Hänen tulee ymmärtää kysymykset (yksiselitteisesti), tietää

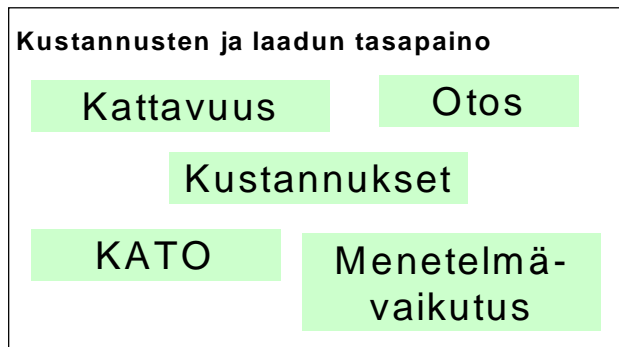
vastaus sekä osata kirjata se lomakkeelle. Tämä asettaa haasteita kysymysten ja ohjeiden muotoilulle sekä lomakkeiden tekniselle rakenteelle.

Kun tiedonkeruumenetelmää valitaan, tulee olla tietoinen tiedonkeruutavan mahdollisista vaikutuksista tutkimustuloksiin. Erityisesti, jos lomakkeella on sensitiivisiä kysymyksiä tai vastausskaaloja, on tiedonkeruumenetelmän osoitettu vaikuttavan saatuihin tuloksiin (Gmel 2000; Fricker et al. 2005; Wang et al. 2005; Dillman et al. 2009).

Toisaalta on myös joukko tutkimuksia ja tutkijoita, jotka ovat nostaneet esille ajatusta, etteivät menetelmäerot olekaan niin suuria kuin aikaisemmin ehkä on luultu (De Leeuw 2005 & 2008; Brener et al. 2006; De Leeuw et al. 2008, 300; Voogt & Saris 2005) tai ettei niitä juuri ole (Greenfield et al. 2000). Menetelmävaikutukset ovat aina kulttuuri- ja aikasidonnaisia (Pohjanpää 2009b).

Kadon kasvaessa ja tiedonkeruukustannusten huiman nousun seurauksena ovat tutkijat pakotettu kompromisseihin mahdollisten menetelmäerojen, kadon ja kustannusten välillä (De Leeuw 2007 & 2008, kuvio 2.4).

**Kuvio 2.4** Kustannusten ja laadun tasapaino tiedonkeruuprosessissa



Lähde: De Leeuw 2007.

### 2.4.3 Mixed mode - eli monimenetelmällisyys

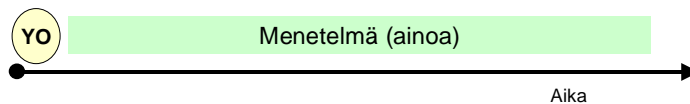
Kun tiedonkeruukustannukset kasvavat ja samanaikaisesti myös osallistuminen tutkimuksiin heikkenee (eli vastausosuudet pienenevät), tulee etsiä uusia keinoja. Menetelmien yhdistämien, monimenetelmällisyys (*mixed mode*) on yksi ratkaisu.

Monimenetelmällisyys tarkoittaa useamman eri tiedonkeruumenetelmän käyttämistä tiedonkeruuprosessissa. Se voidaan toteuttaa monella tapaa (kuvio 2.5). Ensinnäkin

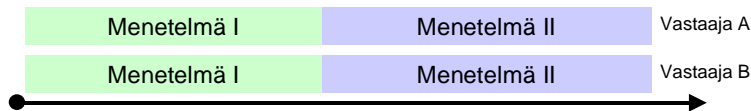
(1) ensimmäinen kontakti tai muistutus (karhuaminen) voidaan tehdä eri menetelmällä kuin varsinainen tiedonkeruu. Toiseksi (2) yksittäisessä haastattelussa voidaan käyttää useampaa eri menetelmää siten, että esimerkiksi osaan (käynti)haastattelun kysymyksistä pyydetään vastaus paperilomakkeella. Tiedonkeruu voidaan toteuttaa myös siten, että (3) eri henkilöiden haastatteluissa käytetään eri menetelmiä. Esimerkiksi eri väestöryhmille tarjotaan eri menetelmiä tai vastaaja saa itse valita useammasta vaihtoehdosta itselleen mieluisimman tiedonkeruumenetelmän. Lisäksi monimenetelmällisyys voidaan toteuttaa siten, että (4) samoille vastaajille käytetään haastattelun eri vaiheissa eri menetelmää. (De Leeuw 2005 & 2007; Dillman et al. 2009.)

**Kuvio 2.5** Monimenetelmällisyyden muotoja tiedonkeruuprosessissa.

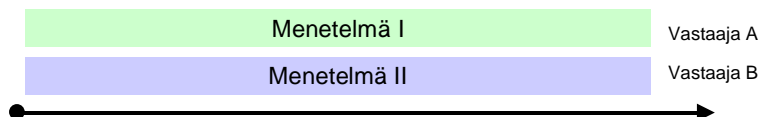
Yhteydenotto eri menetelmällä (multi-mode)



Eri menetelmät sarjassa (mixed mode)



Eri menetelmät rinnan (mixed mode)



Tiedonkeruuprosessin vaikeusaste lisääntyy

Lähde: Tim Macer 2003.

Erityyppiset tiedonkeruutavat mahdollistavat varsin erilaisen otoskoon, mikä vaikuttaa suoraan tietojen yleistettävyyteen. Kalleinta tiedonkeruu on käyntihaastattelujen avulla tehtynä. Eri tiedonkeruumenetelmiä yhdistämällä voidaan usein oleellisesti pienentää vastauskatoa ja kustannuksia, lisätä tavoitettavuutta tai vähentää haastateltavien tiedonantorasitetta (esim. Brennan & Hoek 1992).

## 2.5 Vastauskato vaikuttaa estimaattorin harhaan ja varianssiin

Otantatutkimuksissa päätelmät perustuvat perusjoukon osajoukkoon ja siksi niihin liittyy epävarmuutta. *Estimaattorin harha (bias)* kuvaa sitä, kuinka lähelle perusjoukon parametria estimaattorin arvo keskimäärin osuu, kun tarkastellaan kaikkia mahdollisia tietyllä otantamenetelmällä poimittavia otoksia. *Estimaattorin varianssi (variance)* kuvaa estimaattorin lukuarvojen vaihtelua yli otosten. Otantateoriasta johdetut estimaattorien ominaisuudet perustuvat olettamukseen, että kaikilta otosyksiköiltä saadaan virheettömästi mitattua kiinnostuksen kohteena olevien ominaisuuksien arvot. Käytännössä näin ei ole asian laita, sillä otantatutkimuksissa esiintyy aina vastauskatoa ja mittausvirheitä.

Vastauskato kasvattaa estimaattorin varianssia ja saattaa tehdä estimaattorista harhaisen. Mikäli vastauskadoksi jäävät tutkimusyksiköt poikkeavat tutkittavan ominaisuuden suhteen niistä tutkimusyksiköistä, jotka vastaavat tutkimukseen, on estimointi harhaista. Jos vastauskato on esimerkiksi suurempi niiden keskuudessa, jotka eivät osallistuneet aikuiskoulutukseen, on saatu estimaatti suomalaisten keskimääräisestä aikuiskoulutukseen osallistumisesta systemaattisesti liian pieni, mikäli kadon vaikutusta ei korjata.

Yleisin kadon korjaamiseen liittyvä menetelmä on tulosten painottaminen (*weighting*), jota käytetään erityisesti systemaattisen vinouman vaikutusten poistamiseen. (Lynn 2008, 54–55.)

Käytännössä vastauskato on usein vinoa eli tutkimuksesta jää pois tiettytyyppiset henkilöt. Tämän takia vastausosuus ei yksinään ole riittävä mittari survey-tutkimuksen laadun arviointiin (Groves 2006; Abraham et al. 2006). Vastaajia ja vastaamattomia tulisi verrata juuri tutkimuksen kannalta keskeisten muuttujien osalta.

Ihmiset osallistuvat tutkimuksiin erilaisin perusteluin ja esimerkiksi ikä, sukupuoli, koulutus sekä asuinalue vaikuttavat vastaamistodennäköisyyteen. Lisäksi ihmiset vastaavat kysymyksiin keskenään eri lailla, joten demografiset taustamuuttujat korreloivat saatujen vastausten kanssa. Kadon aiheuttama harha johtuu juuri tästä. (Groves 2006.)

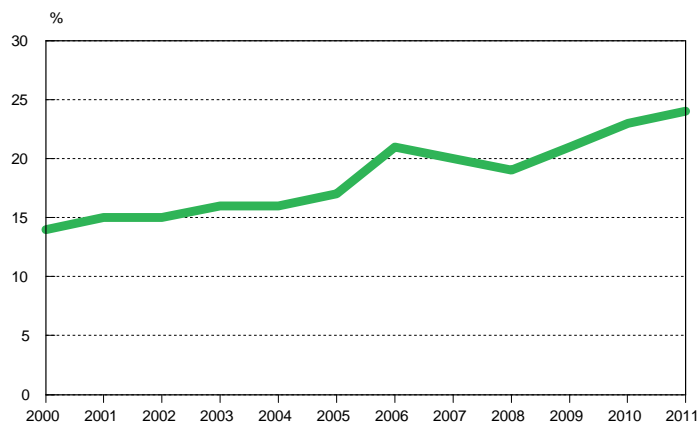
Keinot nostaa vastausosuutta voivat nostaa samalla myös kadon aiheuttamaa virhettä. Etukäteen lähetetyillä saatekirjeillä pyritään lisäämään otoshenkilöiden vastaamishalukkuutta. Monet tutkimukset osoittavat, että palkkiolla saattaa olla vastausosuutta nostava vaikutus, mutta valitettavasti se voi samanaikaisesti myös kasvattaa kadon aiheuttamaa virhettä. Myös haastattelijoiden yhteydenottojen lukumäärä ja/tai haastattelijan omat henkilökohtaiset ominaisuudet (kuten esimerkiksi sukupuoli ja ikä) voivat vaikuttaa katovirheeseen. (Groves 2006.)

Koska vastauskatoon jääneiden vastauksia ei ole käytettävissä, on puutteellisuuden vaikutusta tutkimustuloksiin arvioitu niin sanottujen vastahakoisten vastaajien tai myöhään vastanneiden avulla. Joissakin tutkimuksissa on todettu, että myöhään vastanneet vastaavat eri lailla kuin heti tutkimuskutsun saaneet vastaajat (Lin & Scaeffler 1995; Mariano & Kadade 2001; Lynn et al. 2002; Fricker & Tourangeau 2010).

## 2.6 Vastauskadon kasvaminen survey-tutkimuksien suuri haaste

Vastauskadon kasvaminen on ollut vääjäämätöntä ja lähes pysyvä ilmiö 1900-luvun lopun ja erityisesti 2000-luvun alun survey-tutkimuksissa (kuvio 2.6). Vastamattomuuden syitä on etsitty kaikkialta. Keskeisenä nähdään kansalaisten väsyminen, kun erilaiset tutkimukset, tiedustelut ja ”gallupit” ovat lisääntyneet lähes räjähdysmäisesti. Tutkimusten lisääntyminen on ajoittunut saman aikaan kuin yksilöllisyyden arvostus on kasvanut ja yhteisöllisyys ja yhteisvastuullisuus taas menettäneet merkitystään.

**Kuvio 2.6** Työvoimatutkimuksen (1. rotaatio) vastauskato vuosina 2000–2011, %



Lähde: Tilastokeskus

Kato-osuuden kasvu on yhteydessä myös yksittäisistä kysymyksistä kieltäytymisiin. Esimerkiksi Amerikassa tutkijat (Yan, Curtin & Jans 2010) havaitsivat, että vastaaminen tuloja koskeviin kysymyksiin on vähentynyt. Eräkato tuloja koskevissa kysymyksissä saattaa olla jopa 20–40 prosenttia, kun se muissa survey-tutkimuksen kysymyksissä on keskimäärin yhdestä neljään prosenttia (De Leeuw 1992). Tutkijat uskovat, että yksikkö- ja eräkatoon vaikuttavat mahdollisesti samat mekanismit.

Suomessa eräkato ei ole vielä lisääntynyt. Esimerkiksi Tilastokeskuksen keskeisissä survey-tutkimuksissa ei ole havaittu kovinkaan paljoa puutteellisuutta yksittäisten tutkimuskysymysten osalta, kun tiedot kerätään joko puhelin- tai käyntihaastatteluin. Tältä pohjalta voidaan arvioida, että erä- ja jonkin verran myös yksikkökato ovat kulttuuriin ja aikaan sidottuja ilmiöitä.

Vastauskato jaetaan usein kadon syyn mukaan tavoittamattomiin, kieltäytyneisiin ja muihin syihin. Kadon syyt saattavat poiketa eri aikakausina. Tällä vuosituhanella on muun muassa ollut näkyvissä erityisesti tavoittamattomiksi jääneiden määrän kasvu (Vastauskadon vuosiraportti 2009, Tilastokeskus). Tilastokeskuksen aikuiskoulutustutkimuksen (käyntihaastattelu) katokoodit on esitetty liitteessä 3.

Kadon syiden analysointi voi auttaa kohdistamaan resursseja tehokkaammin: jos tavoittamattomaksi jääminen on erityisen suurta, kannattaa kehitellä keinoja ensikontaktin onnistumiseksi. (Daniel 1975; Groves ja Couper 1998; Groves 2006.)

### 3. Tutkimusaineisto ja kuinka siihen päädyttiin

Aikuiskoulutustutkimuksessa (AKU) selvitetään aikuisväestön koulutukseen osallistumista, oppimista ja osaamista. Lisäksi tilasto tuo tietoa aikuisten opiskelumielipiteistä ja -käsityksistä sekä koulutushalukkuudesta ja -tarpeesta. Tilastokeskus toteuttaa tutkimuksen yhdessä opetus- ja kulttuuriministeriön kanssa.

Ensimmäinen aikuiskoulutustutkimus tehtiin vuonna 1980. Vuodesta 1990 lähtien tutkimus on tehty noin viiden–kuuden vuoden välein, viimeisin vuonna 2006. Seuraavan kerran tietoja kerätään syksyllä 2012.

Tutkimusten kohdejoukkona on Suomessa asuva 18–64-vuotias aikuisväestö. Tutkimustiedot perustuvat noin 5 000 henkilön otoksella toteutettuihin käyntihaastatteluihin. Haastattelu kestää keskimäärin noin tunnin.

Aikuiskoulutustutkimuksen tuloksia on julkaistu Tilastokeskuksen Internet-sivuilla ja vuonna 2008 ilmestyneessä raportissa (Pohjanpää et al. 2008). Tilastokeskus: <http://tilastokeskus.fi/til/aku/index.html>

#### 3.1 Kurkkaus katoon: päätettiin tehdä AKU-laatatutkimus

Aikuiskoulutustutkimuksen vastauskato kasvoi 1990-luvun alle 20 prosentista aina 34 prosenttiin vuonna 2006. Vastaavaa kadon kasvua on ilmennyt myös muissa survey-tutkimuksissa. Vaikka tutkimuksen vastauskato olikin käytettävissä olevien rekisteritietojen perusteella arvioituna melko tasaisesti jakautunut (Pohjanpää et al. 2008, 166–177), syntyi huoli tilastoaineistojen laadusta tutkittavien indikaattorien suhteen. Tämä johti aikuiskoulutustutkimukseen liittyvän laatututkimuksen tekemiseen. Oli aika kurkata katoon.

Vuonna 2007 toteutetulla tutkimuksella haluttiin lisätietoa vastauskatoon jääneistä ja sen haluttiin vastaavan kahteen kysymykseen:

- (1) *Keitä ovat ne henkilöt, joita survey-tutkimuksella ei saada kiinni?*
- (2) *Vastaavatko katoon jääneet samalla tapaa kuin tutkimukseen osallistuneet?*

Laatututkimus suunnattiin niille, jotka eivät vuotta aikaisemmin osallistuneet aikuiskoulutustutkimukseen jääden joko tavoittamattomiksi tai kieltäytyen



osallistumasta vedoten muun muassa aikapulaan. Tutkimuksen toteutus ja otoksen muodostaminen on kuvattu seuraavissa luvuissa.

### 3.3 Laatatutkimuksen otos oli vastauskatoa vuonna 2006

Tutkimuksen tarkoituksena on analysoida Aikuiskoulutustutkimus 2006 -aineiston laatua keräämällä AKU 2006 -aineistoa ”*niin hyvin kuin mahdollista*” vastaava tutkimusaineisto. Laatatutkimuksen otoksena on osa AKU 2006 -tutkimuksessa vastauskatoon jääneistä henkilöistä (kuvio 3.1).

**Kuvio 3.1** AKU 2006:n aineisto ja vastauskato sekä AKU-laatatutkimuksen otos.

AKU 2006 (n=4373) vastanneet		Kato (n=2372)	
Muu maa		Muu maa	
Länsi- ja Etelä-Suomi			Länsi- ja Etelä-Suomi  AKU- LAATU otos (n=1072)

Tutkimus suunnattiin niille, joita ei ollut vuonna 2006 tavoitettu tai jotka kieltäytyivät tuolloin epäsuorasti tai ajanpuutteeseen vedoten. Tutkimuksen ulkopuolelle rajattiin esimerkiksi kaikki periaatteellisesta syystä tai tutkimusaiheen takia kieltäytyneet. Näin siksi, että haluttiin kunnioittaa tiukan kieltäytymisen ilmaisseita henkilöitä eikä haastattelijoita haluttu asettaa hankalaan tilanteeseen. AKU 2006 katokoodit ja niiden yhteys laatatutkimuksen otokseen on esitetty liitteessä 3.

Tutkimus rajattiin Etelä- ja Länsi-Suomen lääneihin aineiston analysoinnin helpottamiseksi. AKU 2006 -tutkimuksessa Etelä-Suomen läänissä oli suurin vastauskato (41 prosenttia). Länsi-Suomen läänissä vastauskato oli samaa tasoa kuin

AKU 2006:ssa keskimäärin eli noin 34–35 prosenttia. Valitut läänit muodostavat yhdessä noin 75–77 prosenttia koko alkuperäisen AKU 2006:n otoksesta ja saadusta aineistosta. AKU 2006:n kokonaiskadosta Etelä- ja Länsi-Suomen läänien vastauskato kattaa 82 prosenttia.

Katokoodien kiinnittämisen myötä tuli laatututkimukseen valituksi Länsi-Suomen läänissä 53 ja Etelä-Suomen läänissä 56 prosenttia kyseisen läänin kadosta (taulukko 3.1). Koko AKU 2006 kadosta nimetyt kadon syyt ovat 54 prosenttia.

Etelä-Suomen läänissä laatututkimuksen otoksessa on eniten vuonna 2006 tavoittamattomia, kun taas Länsi-Suomen otoksessa on melko tasaisesti ajanpuutteen takia kieltäytyneitä ja tavoittamattomaksi jääneitä.

**Taulukko 3.1** Aikuiskoulutustutkimus 2006:n vastauskato kadon syiden mukaan sekä tutkimukseen vastanneet koko aineistossa sekä Länsi- ja Etelä-Suomen lääneissä, lukumäärät ja %-osuudet. AKU-laatututkimuksen otos on merkitty taulukkoon vihreällä (oikea yläkulma).

<b>Kadon syy vuonna 2006</b>	<b>Koko aineisto</b>		<b>Länsi-Suomen lääni</b>		<b>Etelä-Suomen lääni</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Ei tavoitettu	616	9,1	172	7,3	357	12,5
Kieltäytyi, ajanpuute	515	7,6	195	8,3	240	8,4
Epäsuora kieltäytyminen	144	2,1	55	2,3	53	9,1
<b><i>Yhteensä</i></b>	<b>1 275</b>	<b>18,9</b>	<b>422</b>	<b>17,9</b>	<b>650</b>	<b>22,8</b>
Muu vastauskato	1 097	16	371	16	453	18
<b><i>Vastauskato yhteensä</i></b>	<b>2 372</b>	<b>35</b>	<b>793</b>	<b>34</b>	<b>1 160</b>	<b>41</b>
Haastateltu (AKU 2006)	4 373	65	1 562	66	1 695	59
<b><i>Yhteensä</i></b>	<b>6 745</b>	<b>100</b>	<b>2 355</b>	<b>100</b>	<b>2 855</b>	<b>100</b>

AKU-laatututkimuksen otos on suuruudeltaan 1 072 henkilöä. Kolme viidestä laatututkimukseen valitusta tuli Etelä-Suomen läänistä (650 henkilöä) ja loput 40 prosenttia Länsi-Suomen läänistä (422 henkilöä). Laatututkimuksen otoksessa korostui Etelä-Suomi, sillä AKU 2006:n otoksessa vastaava suhde oli 54–46 prosenttia.

Joka toinen otokseen valittu (yhteensä 543 henkilöä) oli kieltäytynyt AKU 2006 - tutkimuksesta vuotta aikaisemmin ja joka toinen kuului niiden joukkoon, joita ei syystä tai toisesta onnistuttu tuolloin tavoittamaan. Kieltäytyneiden osuus oli Länsi-

Suomen läänissä (59 prosenttia) suurempi kuin Etelä-Suomen läänissä (45 prosenttia).

Hiukan useampi kuin joka toinen laatututkimuksen valittu oli mies (54 prosenttia otoksesta miehiä). Miehiä ja naisia oli otoksessa valikoitumissyiden (kieltäytyminen, tavoittamattomuus) suhteen kutakuinkin samassa määrin, eikä otokseen valikoitumissyyt ollut riippuvaisia henkilön iästä (taulukko 3.2). Laatututkimuksen otoksen ja AKU 2006 kadon jakaumien tarkempi vertailu on esitetty luvussa 4.3.

**Taulukko 3.2** AKU-laatututkimuksen otos AKU 2006 kadon syiden sekä henkilön asuinläänin ja sukupuolen mukaan, %-osuudet.

	Länsi-Suomen lääni			Etelä-Suomen lääni		
<i>Kadon syy vuonna 2006</i>	<b>Yht.</b>	<b>Miehet</b>	<b>Naiset</b>	<b>Yht.</b>	<b>Miehet</b>	<b>Naiset</b>
Ei tavoiteta	41	42	40	55	56	54
Kieltäytyy, ajanpuute	46	44	49	37	35	39
Epäsuora kieltäytyminen	13	14	12	8	9	7
<b><i>Yhteensä</i></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
(n)	(422)	(240)	(182)	(650)	(342)	(308)

### 3.3 Tiedonkeruun kulku

Laatututkimus päätettiin toteuttaa käyttämällä useampia tiedonkeruutapoja (mixed mode -tiedonkeruu). Näin siksi, että haluttiin saada edes jotain tietoa vastaamattomilta huomioiden kuitenkin, että resurssit olivat rajalliset. Koettiin myös eettisesti hankalaksi painostaa kovin suuresti jo kertaalleen (vuosi sitten) kieltäytyneitä osallistumaan tutkimukseen ja siksi heille haluttiin tarjota vaihtoehtoisia vastaamistapoja.

Laatututkimuksen tiedonkeruu toteutettiin maaliskesäkuussa 2007, tasan vuotta myöhemmin aikuiskoulutustutkimuksen tiedonkeruuta. Haastateltaville lähetettiin ensin tutkimusta koskeva saatekirje (liite 4), kuten Tilastokeskuksen survey-tutkimuksissa on tapana. Saatekirjeessä ei ollut mainintaa eri tiedonkeruutavoista, vaan oletuksena oli käyntihaastattelu. Kirjeen jälkeen Tilastokeskuksen koulutettu ammattihaastattelija soitti ja yritti saada haastateltavia mukaan AKU 2006:n tapaiseen (pitkään) käyntihaastatteluun. Jos tämä ei tuntunut onnistuvan, ehdotti

haastattelija lyhyttä, puhelimessa tehtävää haastattelua. Jos haastateltava kieltäytyi siitäkin, haastattelija pyrki esittämään muutamana niin sanotun katokysymyksen heti ensimmäisen onnistuneen yhteydenoton yhteydessä. (Kuvio 3.2.)

**Pitkä haastattelu**<sup>6</sup> tehtiin käyntihaastatteluna ja sen pituus oli keskimäärin noin 30 minuuttia. Haastattelu käsitti lähes kaikki (noin 75 prosenttia) Aikuiskoulutustutkimus 2006:n kansallisista kysymyksistä. AKU 2006 haastattelun keskipituus oli noin 55 minuuttia.

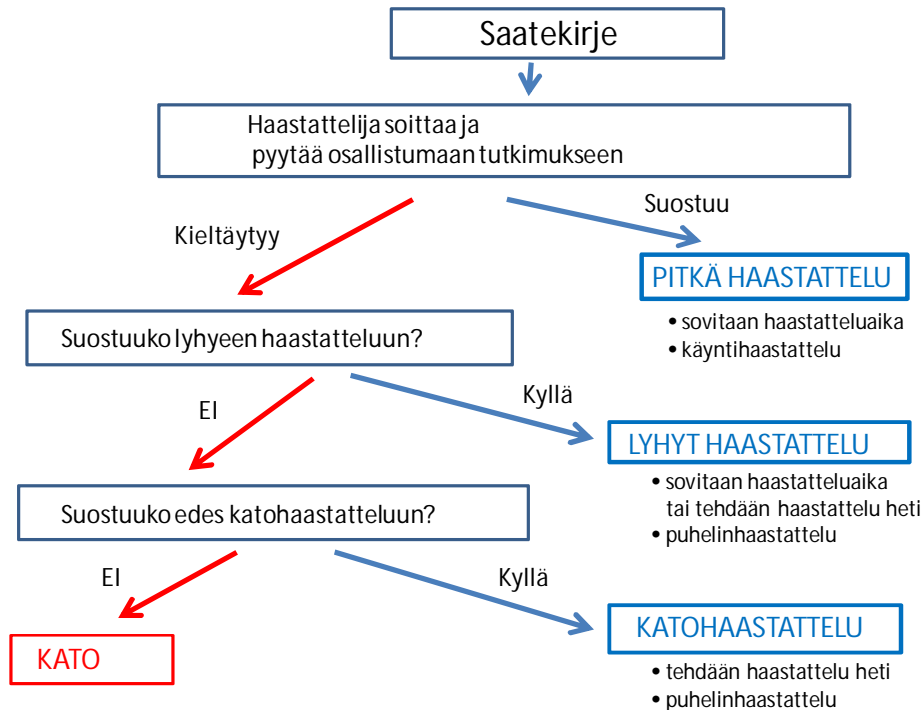
**Lyhyt haastattelu**<sup>1</sup> tehtiin puhelimitse. Lyhyeen haastatteluun oli poimittu joitakin AKU 2006 -tutkimuksen keskeisiä kysymyksiä. Lähinnä kartoitettiin koulutukseen osallistumista viimeksi kuluneiden 12 kuukauden aikana (kysymys B1), viimeisimmän koulutuksen sisältöä ja volyymia sekä koulutuksen järjestäjää. Lisäksi kysyttiin joitakin koulutukseen liittyviä asennekysymyksiä. Lyhyen haastattelun kesto oli keskimäärin noin 8 minuuttia. Haastattelu voitiin tehdä välittömästi ensikontaktin yhteydessä tai sen tekemiselle oli mahdollista sopia uusi aika.

Ns. **katohaastattelu** tehtiin, jos haastateltava ei suostunut pitkään eikä lyhyeen haastatteluun, mutta suostui vastaamaan noin kahdeksaan katokysymykseen (katso edellä). Katohaastattelu tehtiin puhelimessa välittömästi ensimmäisen kontaktin yhteydessä.

Kaikille pitkään tai lyhyeen haastatteluun suostuneille esitettiin myös 10–17 katokysymystä, jotka koskivat tutkimukseen osallistumista ja Tilastokeskuksen tuttuutta. Katokysymykset on esitetty liitteessä 1.

---

<sup>6</sup> Tutkimuksessa päädyttiin käyttämään nimityksiä *pitkä* ja *lyhyt haastattelu* käynti- ja puhelinhaastattelun sijaan, koska AKU 2006 tapaisen käyntihaastattelun rinnalle haluttiin ensisijaisesti tarjota myös lyhyempi, vain keskeisimmät kysymykset kattava vaihtoehto (*”saadaan edes jotain tietoa”*). Rinnakkainen haastattelumenetelmä (puhelinhaastattelu) oli täten vain ”seuraus” haastattelun kestosta (5–10 min).

**Kuvio 3.2** AKU-laatus tutkimuksen tiedonkeruun rakenne

### 3.4 Tiedonkeruu ja haastattelijatyö

AKU-laatus tutkimuksen tiedonkeruu ajoittui maaliskesäkuulle 2007 eli tasan vuosi aikuiskoulutustutkimuksen jälkeen. Tiedonkeruussa oli mukana 38 Tilastokeskuksen tilastohaastattelijaa (22 haastattelijaa Etelä-Suomen ja 16 Länsi-Suomen läänistä). Haastatteli joille järjestettiin puolen päivän pituinen koulutus projektia varten. Haastattelijat tekivät kaiken kaikkiaan noin 1 800 työtuntia aineiston keräämiseksi.

Haastatteli joilla oli keskimäärin 28–29 henkilöä haastateltavana. Tavoitteena oli, etteivät haastattelijat saaneet samoja henkilöitä haastateltaviksi kuin vuonna 2006. Tässä onnistuttiin suhteellisen hyvin. Vain noin 12 prosenttia (yhteensä 125 otokseen kuuluvaa) kohteista oli haastattelijoiden ”omia” ja loput (88 prosenttia, 906 henkilöä) onnistuttiin vaihtamaan. Saaduista pitkistä haastatteluista 5 prosenttia oli haastattelijan ”omia” kohteita, muista haastatteluista ja vastauskadosta oli AKU 2006:n kanssa samoja kohteita 11–14 prosenttia.

Saatuja haastatteluja kertyi keskimäärin 16 kappaletta haastattelijaa kohden. AKU 2006:ssa haastatteluja kertyi keskimäärin 33 kappaletta kullekin haastattelijalle. Kohteitakin heillä oli tuolloin selvästi enemmän, noin 45.

### 3.5 Käytetyt menetelmät

Tutkimuksessa on käytetty kahta aineistoa. Laatututkimuksen aineisto on vastauskadon ja tutkimusaineiston laatuarviointiin tarkoitettu tutkimusaineisto, minkä lisäksi on käytetty Aikuiskoulutustutkimus 2006:n tilastoaineistoa. Tulokset esitetään *yksinkertaisina frekvenssijakaumina*. Tulokset on esitetty tässä raportissa vakioimattomina, mutta lisävarmistuksena osa tuloksista vakioitiin iän ja sukupuolen suhteen. Tässä käytettiin suoraa vakiointia, joka tehtiin excel-taulukoinnin avulla. Vakioinnin vaikutus tuloksiin on mainittu tekstissä.

Aikuiskoulutustutkimus 2006:n tulokset on esitetty Etelä- ja Länsi-Suomen läänien osalta. Tuloksissa on käytetty vastauskatoa korvaavaa painotusta. Painotuksessa katoon jääneiden ja tutkimukseen osallistuneiden suhteita tarkastamalla saadaan kerätty tutkimusaineisto edustamaan mahdollisimman hyvin perusjoukkoa eli tässä tapauksessa Suomessa asuvaa 18–64-vuotiasta väestöä (pois lukien laitospöytä) Etelä- ja Länsi-Suomen lääneissä. Korostuskertoimet korottavat otostutkimuksessa saadut vastaukset perusjoukkoon siten, että yhden henkilön antamat vastaukset edustavat lopullisessa aineistossa noin 300 henkilöä (sama ikä, sukupuoli ja asuinalue).

Eri tiedonkeruumenetelmällä saatujen tulosten eroja (riippuvuutta) on tarkasteltu tilastollisen merkitsevyyden osalta  *$\chi^2$ -riippumattomuustestin* avulla.

$\chi^2$ -riippumattomuustestin avulla voidaan päätellä, johtuuko muuttujien välillä oleva riippuvuus sattumasta vai voidaanko riippuvuutta pitää tilastollisesti merkitsevä. Testaus perustuu (reunajakaumista saatuihin) teoreettisiin eli odotettuihin (*expected*) frekvensseihin ( $f_e$ ), joita verrataan saatuihin (havaitut) (*observed*) frekvensseihin ( $f_o$ ).  $\chi^2$ -testisuure on sitä suurempi, mitä enemmän odotetut ja havaitut frekvenssit poikkeavat toisistaan. Testisuureen ja vapausasteiden perusteella voidaan määritellä merkitsevyystaso, joka kertoo, kuinka suuri riski on, että riippuvuus johtuu sattumasta.

$\chi^2$ -testisuure lasketaan kaavalla:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}, \quad (3.0)$$

missä  $f_o$  on havaittu frekvenssi ja  $f_e$  odotettu frekvenssi.

Vapausasteiden määrä  $df$  saadaan kertomalla rivien ja sarakkeiden lukumäärät:

$$df = (r - 1) * (k - 1),$$

missä  $r$  on taulukon rivien ja  $k$  sarakkeiden lukumäärä.

$\chi^2$ -testisuuren ja vapausasteiden avulla saadaan taulukosta (käytännössä tilasto-ohjelmisto tulostaa) lukuparia vastaavat p-arvot (todennäköisyydet).

Tässä raportissa raportoidaan merkitsevyystasot tulostaulukoiden ja -kuvioiden alaviitteissä kolmen yleisesti käytetyn riskitason avulla:

***	p<0,001
**	p<0,01
*	p<0,05

$\chi^2$ -riippumattomuustestin testisuureet laskettiin SAS-ohjelmistolla.

### *Logistinen regressio*

Kolmantena analyysimenetelmänä on käytetty **logistista regressiomallitusta**. Siinä on tarkoituksena selittää frekvenssien vaihtelua erilaisten mallien avulla. Logistisen regressiomallissa, jossa on vain yksi selitettävä muuttuja, kutsutaan **logit-mallitukseksi**. Logistinen regressiomalli on periaatteeltaan sama kuin lineaarinen regressioanalyysi, erona on se, että selitettävä muuttuja ei ole jatkuva vaan dikotominen eli kaksiluokkainen (esimerkiksi 1 = *osallistuu tutkimukseen*, 0 = *ei osallistu tutkimukseen*). Mallissa selittävät muuttujat voivat olla joko jatkuvia tai diskreettejä (epäjatkuvia, kategorisia, luokittelevia).

Logistisen regressioanalyysin etuna on se, ettei se aseta tutkivalle ilmiölle jakaumaoletuksia.

Selitettävä muuttuja on  $R_i$ :

$$R_i = \begin{cases} 1, & \text{jos } i \text{ osallistui tutkimukseen (tai aikuiskoulutukseen)} \\ 0 & \text{muulloin} \end{cases} \quad (3.1)$$

Merkitään todennäköisyyttä yksikölle  $i$  osallistua tutkimukseen  $\pi_i$ :llä. Tällöin

$$P(R_i = 1) = \pi_i \quad (3.2)$$

Vastaavasti ei-osallistumisen todennäköisyyttä kuvaa kaava:

$$P(R_i = 0) = 1 - \pi_i \quad (3.3)$$

koska  $P(R_i = 1) + P(R_i = 0) = 1$  kaikilla  $i$ :n arvoilla eli todennäköisyys, että havainto (yksilö) osallistuu tutkimukseen tai ei osallistu on tasan yksi. Tällöin on kyse niin sanotusta Bernoullin kokeesta ja satunnaismuuttuja  $R$  noudattaa Bernoullin jakaumaa  $R \sim \text{Ber}(\pi)$ , missä  $0 \leq \pi \leq 1$ .

Kuvaamalla  $r$ :lla havaittua arvoa satunnaismuuttujalle  $R$ , missä  $r$  on binomijakautunut, voidaan  $R$ :n todennäköisyysjakauma kuvata Bernoullin jakaumana:

$$P(R = r) = \pi^r (1 - \pi)^{1-r} \quad (3.4)$$

Jos voimme olettaa, että jokaisella yksiköllä on sama todennäköisyys  $\pi$  osallistua tutkimukseen, niin jokainen satunnaismuuttuja  $R_1, R_2, \dots, R_n$  toteuttaa Bernoullin yhtälön:

$$P(R_i = r_i) = \pi^{r_i} (1 - \pi)^{1-r_i} \quad (3.5)$$

jossa  $r_i$  on havaittu kaksiluokkainen vastaamista kuvaava muuttuja  $i$ :nnelle yksikölle. Yksiköitä on  $n$  kappaletta,  $i = 1, 2, \dots, n$ .

Tällöin  $R_i$  :n odotusarvo on

$$E(R_i) = \mu_i = \pi_i \quad (3.6)$$

ja varianssi

$$\text{Var}(R_i) = \sigma_i^2 = \pi_i(1 - \pi_i) \quad (3.7)$$

Oletetaan nyt, että aineisto voidaan jakaa tutkittavien muuttujien mukaan  $k$  kappaletta luokkia (esimerkiksi jakamalla iän, sukupuolen ja alueen suhteen saadaan luokituksesta riippuen vaikkapa  $4 * 2 * 2 = 16$  toisensa poissulkevaa luokkaa). Merkitään  $n_i$  havaintojen lukumäärää luokassa  $i$  ja  $y_i$ :llä niiden yksiköiden (henkilöiden) määrää, joilla on tarkastelun kohteena olevan ilmiön ominaisuuksia (esimerkiksi osallistuvat tutkimukseen) ryhmässä  $i$ .

$y_i$  = tutkimukseen osallistuneet ryhmässä  $i$

Nyt  $y_i$  on satunnaismuuttujan  $Y_i$  havaittu arvo, joka saa arvoja  $0, 1, \dots, n_i$ . Jos kaikki  $n_i$  havaintoa kussakin ryhmässä on toisistaan riippumattomia ja jos niillä kaikilla on



sama todennäköisyys  $\pi_i$  suhteessa tutkittavaan ilmiöön (esimerkiksi todennäköisyys osallistua tutkimukseen), niin tällöin  $Y_i$  on binomijakautunut parametrein  $\pi_i$  ja  $n_i$  eli

$$Y_i \sim B(n_i, \pi_i)$$

Satunnaismuuttujan  $Y_i$  todennäköisyysjakauma saadaan kaavasta

$$P\{Y_i = y_i\} = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (3.8)$$

kun  $y_i = 0, 1, \dots, n_i$ . Tässä  $\pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}$  on todennäköisyys, kun  $y_i$  toteutuu. Kerroin kuvaa sitä, kuinka monta kertaa  $y_i$  toteutuu  $n_i$  määrässä kokeita.

Odotusarvo ja variassi  $Y_i$ :lle ovat

$$E(Y_i) = \mu_i = n_i \pi_i \quad (3.9)$$

$$Var(Y_i) = \sigma_i^2 = n_i \pi_i (1 - \pi_i) \quad (3.10)$$

Edellä esitetystä luokitellun aineiston analyysistä ”päästään” yksikköaineistoon matemaattisesti ajatellen siten, että ryhmien kooksi määrätään yksi eli ollaan tilanteessa, jossa meillä on  $n$  kappaletta ryhmiä  $k$  eli  $k = n$  ja  $n_i = 1$  kaikilla  $i$ .

Samoin edellä esitetty (kaavat 3.8–3.10) mahdollistaa erikoistapauksen, jossa todennäköisyys on sama kaikille yksiköille  $i$  ja meillä on vain yksi ryhmä eli  $k = 1$  ja  $n_1 = n$ .

Näin ollen, jos selitettävä muuttuja on diskreetti ja selittävät muuttujat keskenään riippumattomia, voimme käyttää Bernoullin jakaumaa yksikköaineistossa ja binomijakaumaa luokitellussa aineistossa. Molemmat jakaumat johtavat samaan todennäköisyysjakaumaan ja myös estimaatit ja keskihajonta ovat samoja.

Oletetaan, että todennäköisyys  $\pi_i$  on kovariaattien  $X_i$  lineaarikombinaatio

$$\pi_i = X_i' \beta \quad (3.11)$$

jossa  $\beta$  on regressiokerroinvektori.

Kaavaa 3.11 kutsutaan myös lineaariseksi todennäköisyysmalliksi (*linear probability model*). Aineistosta todennäköisyys estimoidaan käyttäen pienimmän neliösumman menetelmää (*ordinary least squares*, OLS).

Ongelmana on, että yhtälön 3.11 vasen puoli (todennäköisyys) saa arvoja väliltä  $(0,1)$ , mutta yhtälön oikea puoli voi saada mitä tahansa arvoja.

Yksinkertainen ratkaisu edellä esitettyyn ongelmaan on muuttaa yhtälöä. Ensinnäkin siirrymme todennäköisyydestä  $\pi_i$  sen *odds*:iin eli vedonlyöntisuhteeseen

$$odds_i = \frac{\pi_i}{1-\pi_i} \quad (3.12)$$

*Odds* on siis todennäköisyyden ja sen ”komplementtitodennäköisyyden” osamäärä tai suosiollisten (*osallistuu tutkimukseen*) ja ei-suosiollisten (*ei osallistu tutkimukseen*) todennäköisyyksien osamäärä.

Seuraavaksi otamme tästä logaritmit eli laskemme *logit*:it (joskus puhutaan myös log-oddseista):

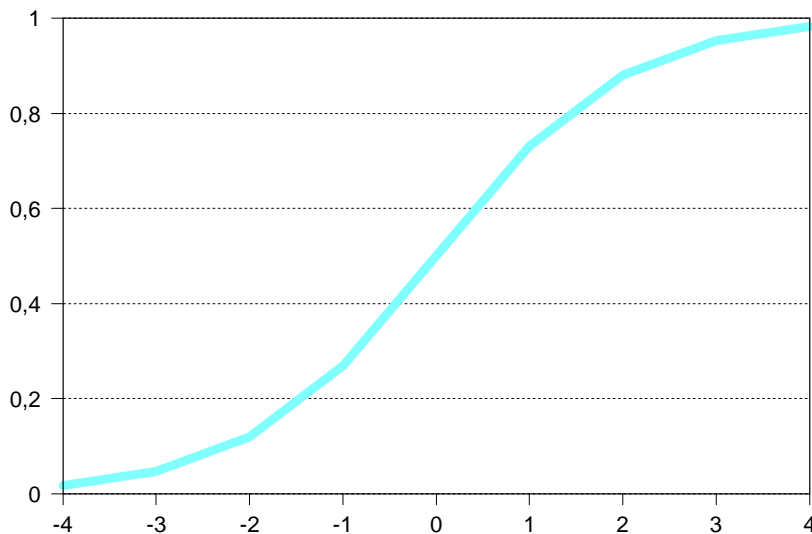
$$g(x_i) = \text{logit}(\pi_i) = \log \left[ \frac{\pi_i}{1-\pi_i} \right] \quad (3.13)$$

Kun yhtälöstä 3.13 ratkaistaan  $\pi_i$ , saadaan

$$\pi_i = \text{logit}^{-1}(g(x_i)) = \frac{e^{g(x_i)}}{1+e^{g(x_i)}} \quad (3.14)$$

Alla olevassa kuviossa (kuvio 3.3) on esitetty logit-muunnos. Kuvio havainnollistaa hyvin arvoaluemuutosta: y-akselilla todennäköisyyden arvoalue  $(0,1)$  ja x-akselilla uusi arvoalue  $(-\infty, +\infty)$ .

**Kuvio 3.3** Logit-muunnos eli logistinen S-käyrä



Nyt olemme valmiit siirtymään logistiseen regressiomallitukseen. Oletetaan, että meillä on  $k$  kappaletta riippumatonta havaintoa  $y_1, \dots, y_k$ . Lisäksi oletamme, että  $i$ :s havainto toteuttaa satunnaismuuttujan  $Y_i$  ja että  $Y_i$  on binomijakautunut

$$Y_i \sim B(n_i, \pi_i) \quad (3.15)$$

Yksikköaineistossa  $n_i = 1$  kaikilla  $i$ :n arvoilla.

Oletamme myös, että todennäköisyyden  $\pi_i$  logit on lineaarikuvaus (kuten 3.11)

$$\text{logit}(\pi_i) = X_i' \beta \quad (3.16)$$

missä  $X_i$  kuvaa vaihtelua (vektori) ja  $\beta$  on regressiokerroin (vektori).

Yhtälöt 3.15 ja 3.16 ovat yleisen lineaarisen mallin (*generalized linear model*) erikoistapauksia, kun satunnaismuuttuja on binomijakautunut ja kun linkkifunktiona on logit-funktio.

Korottamalla yhtälön 3.16 molemmat puolet eksponettiin (ja huomioimalla kaavan 3.13), pääsemme yhtälöön

$$\frac{\pi_i}{1 - \pi_i} = \exp \{X_i' \beta\} \quad (3.17)$$

Kun tästä ratkaistaan  $\pi_i$ , saadaan

$$\pi_i = \frac{\exp \{X_i' \beta\}}{1 + \exp \{X_i' \beta\}} \quad (3.18)$$

Merkitään nyt  $\pi(x_i)$ :llä todennäköisyyttä, että yksittäinen otoshenkilö  $i$  osallistuu tutkimukseen. Jos tämä todennäköisyys riippuu  $p$  kappaleesta selittäviä muuttujia  $X_1, X_2, \dots, X_p$ , niin todennäköisyys voidaan kirjoittaa muotoon:

$$\pi(x_i) = E(R_i | x_i) = P\{R_i = 1 | x_i\} = \frac{e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}}} \quad (3.19)$$

Tämä voidaan esittää myös muodossa:

$$g(x_i) = \log \left[ \frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (3.20)$$

Tässä  $x_{1i}, x_{2i}, \dots, x_{pi}$  ovat muuttujien  $X_1, X_2, \dots, X_p$  arvoja  $i$ :nnelle yksikölle ja kertoimet  $\beta_0, \beta_1, \dots, \beta_p$  ovat tuntemattomia. Tällöin  $g(x_i)$  kuvaa osallistumisen oddsien logaritmia. Linkkifunktio  $g(x_i)$  voi saada mitä tahansa arvoja ja  $\pi(x_i)$  saa arvoja väliltä

(0,1). (Hosmer & Lemeshow 2000, Menard 2002, Lehtonen & Pahkinen 2004, Allison 2009.)

### *Parametrien $\beta_j$ estimointi*

Logit-mallituksen keskeinen tehtävä on estimoida vektorin  $\beta$  (esimerkiksi kaavat 3.17–3.18) arvot eli kertoimien  $\beta_0, \dots, \beta_p$  estimointi aineiston avulla. Logistisessa regressioanalyysissä käytetään yleensä suurimman uskottavuuden (*maximum likelihood, ML*) menetelmää. Linearisessa regressiossa käytetty pienemmän neliösumman menetelmä ei sovellu, koska todennäköisyys  $P$  ei ole lineaarisesti riippuvainen selitettävistä muuttujista. Tästä samasta syystä kertoimien  $\beta_0, \dots, \beta_p$  estimointi tehdään iteratiivisesti.

ML-menetelmässä aineiston perusteella estimoidaan parametrit  $\beta$  siten, että estimaatit antavan aineiston saaminen olisi mahdollisimman todennäköistä. Estimoinnissa käytetään apuna niin sanottua *uskottavuusfunktiota* (*likelihood*)  $L(\theta)$

$$L = \prod_{i=1}^n \left( \frac{p_i}{1-p_i} \right)^{y_i} (1 - p_i) \quad (3.21)$$

Suurimman uskottavuuden menetelmässä valitaan uskottavuusfunktion  $L(\theta)$  parametrit  $\theta_1, \dots, \theta_k$  siten, että  $L(\theta)$  maksimoituu. Usein tämä tapahtuu ratkaisemalla parametrit yhtälöstä

$$\frac{\partial L(\theta)}{\partial \theta_j} = 0 \quad (j = 1, \dots, k) \quad (3.22)$$

Ratkaisussa käytetään tilasto-ohjelmissa usein Newton-Raphson-algoritmia, jolla iteratiivisesti etsitään kyseiset nollakohdat (Allison 2009, 36–39).

### *Vedonlyöntisuhteet, OR*

Logit-mallissa ilmiön todennäköisyyttä tarkastellaan niin sanottujen vedonlyöntisuhteiden (*odds*) logaritmeina (kaava 3.20). Tulokset ilmoitetaan usein vedonlyöntisuhteiden suhteina eli ristitulosuhteina (*odds ratio, OR*):

$$OR = \frac{\pi_i / (1 - \pi_i)}{\pi_1 / (1 - \pi_1)} \quad (3.23)$$

missä  $\pi_i = \pi(x_i)$  on todennäköisyys, että ilmiö vallitsee luokassa  $i$  ja  $\pi_1 = \pi(x_1)$  on todennäköisyys, että ilmiö vallitsee luokassa 1 eli niin sanotussa referenssiluokassa.

Ristitulosuhteiden tulkinnassa saatua  $OR$ :ää verrataan referenssiluokan arvoon (joka siis on 1,00). Tällöin esimerkiksi  $OR = 1,30$  tarkoittaa, että kyseisen ilmiön (esimerkiksi osallistuu tutkimukseen) vetosuhde eli riski, että ilmiö toteutuu tuossa luokassa, on 1,3-kertainen verrattuna riskiin muuttujan ensimmäisessä luokassa (referenssiluokka). (Hosmer & Lemeshow 2000, Menard 2002, Allison 2009.)

### *Mallin hyvyyden arviointi*

Logit-mallituksessa saadun mallin hyvyyttä ja mallituksen onnistumista mitataan usealla eri tavalla. Yleistä on laskea arvio sille, kuinka hyvin malli sopii aineistoon, mikä on mallin selitysosuus tai verrata kahta tai useampaa mallia keskenään sekä arvioida yksittäisten selittäjien tilastollista merkitsevyyttä.

Koska logit-mallitus tapahtuu iteratiivisesti siten, että lähdetään joko täydestä (kaikki selittäjämuuttujat mukana) tai tyhjästä (vain vakiotermi mallissa) mallista, tehdään mallin hyvyyden arviointikin iteratiivisesti. Logit-mallituksessa puhutaan näissä tapauksissa askeltavasta mallituksesta.

Lopullinen malli on saavutettu, kun mallin selityssaste ei enää parane ja kun mallin selittäjät ovat parhaat mahdolliset (parhain selityskyky). Tilastollisessa mallituksessa pyritään aina yksinkertaisimpaan malliin, joka pystyy selittämään tarkasteltavaa ilmiötä.

Logit-mallituksessa saadun mallin hyvyyttä mitataan malliin liittyvällä **Pearsonin  $\chi^2$ -testisuurella**. Tässä on ideana se, että ensin tuotetaan perusmalli, jossa on mukana vain vakiotekijä  $\beta_0$ . Tämän jälkeen muodostetaan malli, jossa on mukana joukko selittäjiä  $\beta_0, \dots, \beta_k$  <sup>7</sup> jossa  $1 \leq k \leq p$  (esimerkiksi kaava 3.32). Saatua malleja verrataan toisiinsa Pearsonin  $\chi^2$ -testin avulla. Jos mallit eivät poikkea toisistaan, niin muuttujien lisääminen malliin ei muuta mallia ja pelkän vakiotermi riittäisi kuvaamaan malliin liittyvää vaihtelua.

Testauksessa on tällöin nollahypoteesina se, että kaikki  $\beta_0, \dots, \beta_k$  ( $1 \leq k \leq p$ ) ovat nollija eli että ennustettujen ja havaittujen arvojen ero ei ole tilastollisesti merkitsevä.

---

<sup>7</sup> Alaindeksin  $k$  käyttöönotto ehdossa perustuu tässä logit-mallituksen iteroivaan luonteeseen. Periaatteessa ehto voitaisiin yhtä hyvin kirjoittaa muotoon  $\beta_0, \dots, \beta_p$  eli käyttäen  $k$ :n sijasta viimeisen  $\beta$ -kertoimen alaindeksinä  $p$ :tä.

Yleisesti Pearsonin  $\chi^2$ -testisuure lasketaan kaavasta

$$\chi^2 = \sum_i \frac{n_i(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i(n_i - \hat{\mu}_i)} \quad (3.24)$$

Kaavan mukaan siis jokainen summan termi on havaittujen ja sovitettujen arvojen  $y_i$  ja  $\hat{\mu}_i$  neliöity erotus, joka jaetaan vielä  $y_i$ :n varianssilla (joka siis on  $\hat{\mu}_i(n_i - \hat{\mu}_i)/n_i$ ).

Pienillä  $p$ :n arvoilla voidaan nollahypoteesi hylätä ja tällöin siis ainakin joku  $\beta_0, \dots, \beta_k$  on erisuuri kuin nolla.

Useimmat tilasto-ohjelmat tulostavat Pearsonin  $\chi^2$ -testisuureen, vapausasteet ja näihin liittyvän  $p$ -arvon. Pienillä  $p$ :n arvoilla saatu malli on tilastollisesti merkitsevä. Nyrkkisääntönä  $\chi^2$ -testisuurelle on perinteisesti pidetty ehtoa  $|\chi^2| \leq 2$ , jotta malli olisi hyvä.

Yksittäisen  $\beta_i$ -kertoimen merkitsevyyttä mitataan käyttäen **Waldin testiä**. Yleisesti testi testaa rajoittamattoman mallin log-uskottavuutta rajoitetun mallin log-uskottavuuteen. Waldin testin nollahypoteesina on, että yksittäinen estimoitu kerroin  $\beta_i$  on nolla.

Käytännössä Waldin testisuure saadaan jakamalla estimoitu  $\beta_i$ -kerroin keskivirheellään ja tämän jälkeen kyseinen osamäärä korotetaan toiseen potenssiin.

Waldin  $\chi^2$ -testisuure kuvataan usein varianssin avulla

$$\chi_{des}^2 = \frac{(\hat{\beta} - \beta_0)^2}{\hat{v}_{des}} \quad (3.25)$$

missä  $\hat{v}_{des}$  on asetelmaan liittyvä varianssiestimaatti  $\hat{\beta}$ :lle. Waldin testisuure on  $\chi^2$ -jakautunut vapausastein  $df = 1$ <sup>8</sup>. (Lehtonen & Pahkinen 2004, 221.)

Tilasto-ohjelmat tulostavat Waldin testisuuretta vastaavan  $p$ -arvon. Suuret testisuureen arvot johtavat nollahypoteesin hylkäämiseen eli toisin sanoen jos

---

<sup>8</sup> Vapausasteiden määrä saadaan kaavasta  $df = m - 1$  eli luokkien määrä miinus yksi, ja koska logit-mallituksessa muuttujat ovat usein dummy-muuttujia, on vapausasteita  $df = 2 - 1 = 1$ .

$p < 0,05$ , on tutkittu kerroin  $\beta_i$  tilastollisesti merkitsevä ja sen pitäminen mukana mallissa on (tilastollisesti)<sup>9</sup> perusteltua.

Logit-mallituksen iteroivasta luoneesta on jo mainittu, ja käytännössä iterointi tapahtuu juuri näiden  $\beta_i$ -kertoimien tilastollista merkitsevyyttä arvioimalla. Tehdään malli, tarkistetaan  $\beta$ -kertoimien tilastollinen merkitsevyys (myös mallin hyvyys on aina syytä tarkastaa), jonka jälkeen malliin lisätään tai siitä poistetaan selittäviä muuttujia. Näin toimitaan, kunnes päädytään parhaaseen mahdolliseen malliin.

Mallin hyvyttä eli sopivuutta aineistoon voidaan tarkastella mallin selitysasteen eli  $R^2$ :n avulla. Koska selittävät muuttujat eivät ole selitettävän muuttujan lineaarikombinaatioita, on perinteisen selitysosuuden  $R^2$  sijaan käytettävä **yleistettyä  $R^2$ :ta** (*generalized  $R^2$* ).

Testi perustuu nollahypoteesiin, että kaikki kertoimet  $\beta_0, \dots, \beta_p$  ovat nollia. Yleistetty  $R^2$  saadaan kaavasta:

$$R^2 = 1 - \exp \left\{ -\frac{L^2}{n} \right\} \quad (3.26)$$

missä  $n$  on otoksen koko ja  $L$  on suurimman uskottavuuden (*maximum likelihood*) estimaatti.

Suurimman uskottavuuden  $L$ -estimaatti lasketaan seuraavasti:

$$L = \prod_{i=1}^n \left( \frac{p_i}{1-p_i} \right)^{y_i} (1 - p_i) \quad (3.27)$$

Kahta mallia voidaan **verrata keskenään** myös tarkastelemalla tilasto-ohjelmien tulostamia  $-2\text{Log}L$ - (jossa  $L$  saadaan kaavasta 3.27), AIC- (*Akaike information criterion*) tai SC-arvoja (*Schwarz criterion*). AIC- ja SC-tunnusluvut saadaan laskemalla  $L$ -arvosta seuraavasti:

$$AIC = -2\text{Log}L + 2k \quad (3.28)$$

$$SC = -2\text{Log}L + k \text{Log } n \quad (3.29)$$

missä  $k$  on estimoitujen parametrien lukumäärä ja  $n$  otoskoko.

---

<sup>9</sup> Jokaisen malliin valitun muuttujan tulee ensisijaisesti olla teoreettisesti perusteltu tutkittavan ilmiön suhteen.

Pelkkä tilastollinen sopivuus ei riitä, ja joskus voi olla perusteltua ottaa malliin mukaan selittäjäksi myös muuttuja, jonka selitysvaikutus tilastollisesti katsottuna ei ole riittävä.

Arvojen pieneneminen tarkoittaa mallin paranemista, ja siten muuttujan lisääminen tai poistaminen mallista on perusteltua.

$-2\text{Log}L$ -arvosta käytetään myös nimitystä poikkeama tai **devianssi** (*deviance*,  $D$ ), jonka avulla voidaan arvioida, kuinka paljon tietty malli  $M$  poikkeaa ainestoon sovitetusta saturoidusta mallista<sup>10</sup>  $M_s$ .

Devianssa voidaan ilmaista seuraavasti

$$D = -2\ln \left( \frac{L(\beta_{tutk})}{L(\beta_{saturoitu})} \right) \quad (3.30)$$

jossa  $L$  saadaan kaavasta 3.29.

Devianssille on myös käyetty merkintää

$$D = D(M_0) - D(M_1) \quad (3.31)$$

jossa

$$M_0 = \log \left( \frac{\pi(x)}{1-\pi(x)} \right) = \text{logit}(\pi(x)) = \beta_0 \quad (3.32)$$

$$M_1 = \log \left( \frac{\pi(x)}{1-\pi(x)} \right) = \text{logit}(\pi(x)) = \beta_0 + \beta_1 x \quad (3.33)$$

(Isotalo 2012, 27).

Devianssi noudattaa asymptoottisesti  $\chi^2$ -jakaumaa vapausastein  $df = k - p$ , missä  $k$  ja  $p$  ovat selittävien muuttujien määrät tutkitussa ja saturoidussa mallissa. Suurilla  $D$ :n arvoilla tarkasteltu malli on huono.

Mallin hyvyttä voidaan mitata myös **Hosmer-Lemeshow -testillä**. Testissä verrataan havaittuja (*observed*) ja ennustettuja (*predected*) frekvenssejä ja testin nollahypoteesina on:

$H_0$ : ei eroja havaittujen ja ennustettujen arvojen välillä

Toisin kuin perus  $\chi^2$ -testissä H-L:n  $\chi^2$ -testisuure saadessa pieniä  $p$ :n arvoja ( $<0,05$ ) jää  $H_0$  voimaan, eikä malli siis tuo mitään uutta tietoa (kertoimet ovat yhtäsuuria kuin täydessä mallissa). Hosmer-Lemeshow -testiä on myös kritisoitu (Allison 2009, 55),

---

<sup>10</sup> Saturoidussa mallissa (täydessä mallissa) on mukana kaikki muuttujat ja niiden väliset yhdysvaikutukset.



eikä se sovi käytettäväksi pienissä aineistoissa. (Hosmer & Lemeshow 2000, Menard 2002, Lehtonen & Pahkinen 2004, Allison 2009.)

### *Logit-mallin käyttö tässä tutkimuksessa*

Tämän tutkimuksen logit-mallitus tehtiin SAS-ohjelmiston proc logistic -proseduurilla, jota varten selittävistä muuttujista tehtiin kaksiluokkaisia dummy-muuttujia. Dummy-muuttujia otetaan malliin selittäjiksi yksi vähemmän kuin alkuperäisessä muuttujassa oli luokkia (eli yksi vähemmän kuin dummy-muuttujia tehtiin).

Logit-mallituksessa lähdettiin niin sanotusta täydestä mallista, jossa on mukana joukko tutkittavan ilmiön kannalta mahdollisia selittäviä muuttujia. Täydestä mallista poistettiin selittäviä muuttujia yksitellen. Poistokriteerinä oli se, ettei muuttuja ollut tilastollisesti merkitsevä (p-arvo suurempi kuin 0,05). Tulososiossa on esitetty lopulliset (tutkijan mielestä parhaat) mallit.

Liitteessä 5 on esitetty logit-mallituksessa käytetyt dummy-muuttujat sekä mallituksien lähtökohta (täydet mallit) ja lopullisten mallien hyvyys (Hosmer ja Lemeshow -testin arvot).

### *3.6 Johdatus tuloksiin*

Kun laatututkimuksen tuloksia verrataan AKU 2006:n tuloksiin, on muistettava kolme seikkaa: tiedot on kerätty (a) *eri ajankohtina (vuoden ero)* ja (b) osittain *eri tiedonkeruumenetelmillä* ja lisäksi saadut (c) *aineistot ovat kooltaan erisuuruisia*. Tiukkojen johtopäätöksiä tekemisessä vertailujen pohjalta on siis syytä olla varovainen. Toisaalta aikuiskoulutukseen liittyvät keskeiset indikaattorit muuttuvat yleensä kovinkin hitaasti, joten tutkimusten välinen aikajänne (yksi vuosi) ei oletettavasti aiheuta suurta ”häiriötä” tuloksiin. Tiedonkeruumenetelmän vaikutuksista tuloksiin on monenlaisia arvioita (De Leeuw 2005 & 2008; Voogt & Saris 2005), eikä vaikutusten suunta ja suuruus ole suinkaan yksiselitteinen. Aineistojen koko vaikuttaa siihen, tulevatko mahdolliset erot tilastollisesti merkitseviksi vai ei – mutta tämä ei kerro mitään erojen merkittävyydestä.

Tämän tutkimuksen tarkoitus ei ole pyrkiä täydelliseen tulkintaan katoon jääneistä tai vastauskadon vaikutuksista haastattelututkimusten laatuun. Tutkimus on yksi kertomus, yksi kuva siitä, mitä tietoa meiltä jäi saamatta vuonna 2006 aikuiskoulutustutkimuksessa. Tiedonkeruumenetelmän vaikutusten analysointi rajataan tämän tutkielman ulkopuolelle.

#### 4. Tulokset: mitä kato kertoo meille?

Seuraavaksi esitellään laatututkimuksen tulokset. Tulososa on jaettu kolmeen: A-osassa esitetään laatututkimuksen vastausosuuksia ja -jakaumia sekä tarkastellaan taustamuuttujien kautta, keitä he ovat, joita emme vuonna 2006 saaneet mukaan tutkimukseen mutta vuonna 2007 saimme (tutkimuskysymys 1, sivu 22). B-osassa paljastetaan, mitä miltä katoon aikaisemmin jääneet ovat survey-tutkimuksista eli esitetään niin sanottujen katokysymysten tulokset (tämäkin liittyy tutkimuskysymykseen 1). C-osassa verrataan katoon jääneiden vastauksia vuoden 2006 aikuiskoulutustutkimuksen tuloksiin eli pohditaan, muistuttavatko katoon jääneet tutkimuksen keskeisten indikaattoreiden osalta osallistuneita (tutkimuskysymys 2).

#### A Tutkimukseen osallistuneet

##### 4.1 Valmiutta osallistua on: vastausosuus 59 prosenttia

Laatututkimuksen vastausosuus oli varsin korkea, sillä 59 prosentilta netto-otokseen kuuluvista saatiin ainakin jotain tietoa. Vähän yli joka kymmenes osallistui AKU 2006:tta vastaavaan käyntihaastatteluun, lähes 40 prosenttia antoi vastaukset lyhyemmässä puhelinhaastattelussa ja harvempi kuin joka kymmenes vastasi vain katokysymyksiin (taulukko 4.1).

**Taulukko 4.1** AKU-laatututkimukseen osallistuneet ja katoon jääneet vuonna 2007

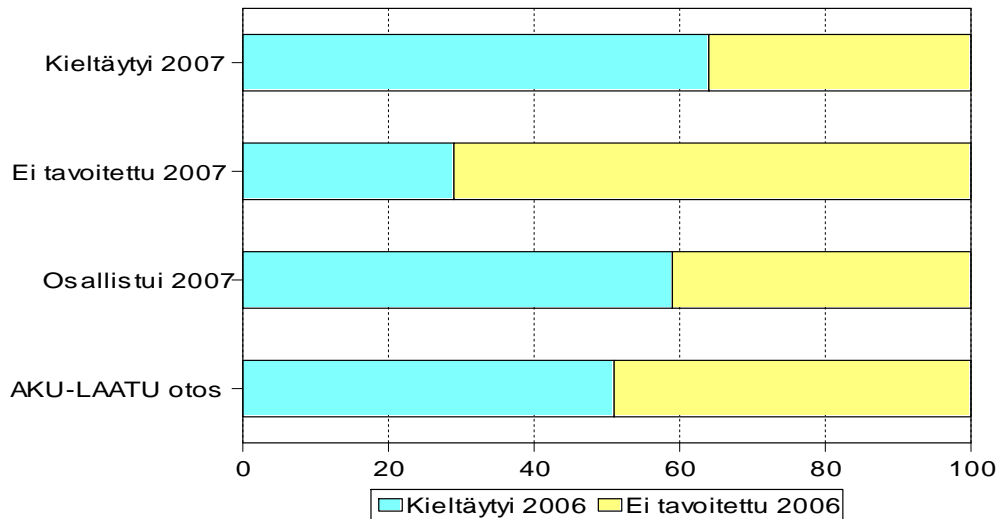
	Lukumäärä	(brutto) %	Kumulaatiivinen (brutto), %	(netto) %	Kumulaatiivinen (netto), %
Pitkä haastattelu	138	13	13	13	13
Lyhyt haastattelu	403	38	50	39	52
Katohaastattelu	70	7	57	7	59
Vastauskato	420	40	96	41	100
Ylipeitto	41	4	100	-	
<b>Yhteensä</b>	<b>1 072</b>	<b>100</b>		<b>100</b>	

Tutkimuksen ylipeitto oli selkeästi suurempi kuin haastattelututkimuksissa yleensä, noin 4 prosenttia. Tulos ei ole kovinkaan kummallinen, kun huomioidaan otoksen määrittelytapa (AKU 2006:n vastauskato). Kaksi kolmesta ylipeittotapauksesta koodattiin rekisteritietojen (otoksen poiminnan) päivityksen yhteydessä. Suurin osa

ylipeitosta muodostui laitoshoidossa pysyvästi olevista (23 henkilöä). Yksitoista henkilöä oli muuttanut pysyvästi ulkomaille ja kuusi henkilöä oli kuollut. Hiukan yllättävää oli, että vain kolmasosa ylipeitosta oli 45 vuotta täyttäneitä. Kahta lukuun ottamatta kaikki vuoden 2007 ylpeittotapaukset (95 prosenttia) olivat vuotta aikaisemmin tavoittamattomia. AKU 2006:ssa ylpeitto oli 0,8 prosenttia.

41 prosenttia AKU 2006:n vastauskadosta kieltäytyi laatututkimuksestakin. Katoon jäämisen syyt olivat usein samat kuin vuonna 2006: kieltäytyneistä 64 prosenttia kieltäytyi toistamiseen ja tavoittamattomista 71 prosenttia oli kadoksissa AKU 2006:n aikoihinkin (kuvio 4.1).

**Kuvio 4.1** Laatututkimuksen vastauskato, tutkimukseen osallistuneet ja tutkimuksen otos AKU 2006 vastauskadon syiden (kieltäytyi, ei tavoitettu) mukaan, %-osuudet.



#### 4.2 Kato ei sittenkään huono-osaisempaa

Laatututkimukseen osallistuminen ei riipu haastateltavan sukupuolesta eikä iästä, mutta sen sijaan asuinlääni, koulutustausta, äidinkieli ja katoon jäämisen syy vuonna 2006 ovat tilastollisesti merkitseviä tai melkein merkitseviä taustamuuttujia tarkasteltaessa tutkimukseen osallistumista. Tarkastelussa huomioidaan vastanneiksi myös ne 70 henkilöä, jotka osallistuivat vain katoa haastatteluun.

Länsi-Suomen läänissä laatututkimuksen vastausosuus oli 66 ja Etelä-Suomessa 55 prosenttia (taulukko 4.2). Aikuiskoulutustutkimuksessa vuonna 2006 länsisuomalaisten vastausosuus oli sama 66 prosenttia ja eteläsuomalaisten 59

prosenttia. Vain Etelä-Suomen läänissä vastattiin laatututkimukseen hiukan huonommin kuin varsinaiseen AKU 2006:een.

Perusasteen koulutuksen saaneet osallistuivat laatututkimukseen harvemmin kuin keski- tai korkea-asteen koulutuksen saaneet. Heistä joka toinen (51 prosenttia) osallistui, kun muissa koulutustaustaryhmissä osallistuneita oli 62 prosenttia. Ero koulutusryhmien välillä koski lähinnä lyhyeen puhelinhaastatteluun osallistumista. (Taulukko 4.3.) Tilanne poikkesi AKU 2006 vastausosuuksista siten, että korkea-asteen koulutuksen saaneet osallistuivat laatututkimukseen harvemmin kuin AKU 2006:een (vastausosuus 71 prosenttia).

**Taulukko 4.2** Laatututkimukseen osallistuminen vuonna 2007 läänin<sup>1</sup> mukaan, %.

	Etelä-Suomi	Länsi-Suomi	Lukumäärä
Pitkä haastattelu	12	15	(138)
Lyhyt haastattelu	37	42	(403)
Katohaastattelu	5	9	(70)
Vastauskato	45	34	(420)
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	
(n)	(623)	(408)	(1 031)

<sup>1</sup> Erot tilastollisesti merkitseviä: \*\*.

**Taulukko 4.3** Laatututkimukseen osallistuminen vuonna 2007 koulutustaustan<sup>1</sup> mukaan, %.

	Perusaste	Keskiaste	Korkea aste
Pitkä haastattelu	14	14	12
Lyhyt haastattelu	30	41	43
Katohaastattelu	7	7	7
Vastauskato	49	38	38
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	<b>100</b>
(n)	(253)	(475)	(303)

<sup>1</sup> Erot tilastollisesti merkitseviä: \*.

Suomea äidinkielenään puhuvista 59 ja ruotsia puhuvista 66 prosenttia osallistui laatututkimukseen. Jos äidinkieli oli jokin muu kuin suomi tai ruotsi, oli osallistumisaste vain 35 prosenttia (tosin heitä oli otoksessakin varsin vähän, 37 henkilöä). AKU 2006:n vastausosuus suomenkielisillä oli 66 ja ruotsinkielisillä 51 prosenttia. Muuta kuin suomea tai ruotsia äidinkielenään puhuvista aikuiskoulutustutkimukseen osallistui joka toinen vuonna 2006. Ruotsinkielisten

vastausosuuksia tarkasteltaessa on huomattava, että laatututkimuksessa ei ole mukana Ahvenanmaata toisin kuin AKU 2006:ssa oli.

Vuonna 2006 kieltäytyneistä 67 prosenttia, mutta tavoittamatta jääneistä vain 51 prosenttia antoi jotain tietoja laatututkimuksessa vuonna 2007. Aikaisemmin kieltäytyneet osallistuivat nyt erityisesti lyhyeen haastatteluun. Sen sijaan tavoittamatta jääneet osallistuivat pitkään haastatteluun useammin kuin vuosi aikaisemmin kieltäytyneet. (Taulukko 4.4.)

**Taulukko 4.4** AKU-laatututkimukseen osallistuminen vuonna 2007 AKU 2006:n katotilanteen<sup>1</sup> mukaan, %.

	Kieltäytyi vuonna 2006	Ei tavoitettu vuonna 2006	Lukumäärä
Pitkä haastattelu	10	17	(138)
Lyhyt haastattelu	47	30	(403)
Katohaastattelu	9	4	(70)
Vastauskato	33	49	(420)
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	
(n)	(541)	(490)	(1 031)

<sup>1</sup> Erot tilastollisesti merkitseviä: \*\*\*.

### 4.3 Jakaumat muistuttavat toisiaan

Vastausosuuksien lisäksi laatututkimukseen osallistuneiden jakaumia on syytä verrata AKU 2006:n jakaumiin ja otokseen. Vertailun tarkoituksena on tarkentaa kuvaa siitä, kuinka edustava saatu laatututkimuksen aineisto on suhteessa AKU 2006:een. Lisäksi tässä luvussa kuvataan laatututkimuksen osallistuneiden työelämään liittyviä taustatietoja sekä Internetin ja vieraiden kielten käyttämistä.

Taulukossa 4.5 on esitetty laatututkimuksen ja alkuperäisen aikuiskoulutustutkimuksen otokset ja vastausjakaumat. Kun verrataan laatututkimuksen otosta (sarake C) laatututkimukseen vastanneisiin (sarake B), huomataan että vastaavuus on lähes täydellisestä muiden taustamuuttujien paitsi asuinläänin suhteen. Vastanneiden joukossa on Etelä-Suomessa asuvia suhteessa vähemmän kuin tutkimuksen otoksessa. Tämä tarkoittaa siis, että vastausosuudet laatututkimuksessa olivat melko samanlaisia eri väestöryhmissä (kuten edellisessä

luvussa todettiin). Poikkeuksena on Etelä-Suomessa asuvien korkea vastaamattomuus – sama ilmiö kuin AKU 2006 -tutkimuksessa.

Kun verrataan laatututkimuksen otosta AKU 2006 vastanneisiin (sarakkeita C ja D), huomataan että laatututkimuksen otoksessa on hiukan yliedustettuna miehet, alle 34-vuotiaat, perusasteen koulutuksen saaneet sekä Etelä-Suomen läänissä asuvat. Nämä ovat ne ryhmät, joissa AKU 2006:ssa oli keskimääräistä suurempi vastauskato. Sen sijaan laatututkimuksen otoksessa on hiukan aliedustettuna suomea äidinkielenään puhuvat, korkea-asteen koulutuksen saaneet ja 55 vuotta täyttäneet. Tämä vertailu kertoo, kuinka hyvin laatututkimuksen otos kuvaa AKU 2006:een vastanneita (ja luo pohjan tutkimusten tulosten keskinäiselle vertailulle). Osittain esille nostetut pienet erot johtuvat AKU 2006 vastauskadon jakaumasta, osittain laatututkimuksen otoksen poimintaperiaatteista.

Kun verrataan laatututkimukseen osallistuneita AKU 2006 osallistuneisiin (sarakkeita B ja D), huomataan että laatututkimuksessa on hiukan yliedustettuna miehet ja Etelä-Suomen läänissä asuvat ja hiukan aliedustettuna suomea äidinkielenään puhuvat. Sen sijaan laatututkimuksessa on selvästi aliedustettuna korkea-asteen koulutuksen saaneet ja 55 vuotta täyttäneet ja yliedustettuina 25–34-vuotiaat. Vastaavanlainen tulos saadaan, kun vertailun kohteen ovat pitkään käyntihaastatteluun osallistuneet vuosina 2007 ja 2006.

Jakaumat AKU 2006:ssa ja laatututkimuksen aineistossa on laskettu ilman painoja. Lisäksi tässä tarkastelussa on AKU 2006:n osalta mukana kaikki läänit, ei siis vain Etelä- ja Länsi-Suomi, kuten laatututkimuksessa (vrt. taulukko 4.6).

**Taulukko 4.5** AKU-laaturutkimuksen otoksen ja vastanneiden<sup>1</sup> jakaumat vuonna 2007 sekä vuoden 2006 aikuiskoulutustutkimuksen otoksen ja vastanneiden jakaumat sukupuolen, iän, koulutusasteen ja alueen mukaan sekä 18–64-vuotias väestö vuoden 2005 lopulla.

	AKU-LAATU		AKU 2006		Väestö 2005		
	vastanneet <sup>1)</sup>		otos	vast.	otos		
	lkm	%	%	%	%	%	tuhatta
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
<b>Sukupuoli</b>							
Mies	320	52,4	54,3	49,5	50,6	50,5	1 672
Nainen	291	47,6	45,7	50,5	49,4	49,5	1 639
<b>Ikä</b>							
18–24-v. <sup>2)</sup>	53	8,7	8,2	5,7	5,9	13,8	458
25–34-v.	229	37,5	37,5	30,2	30,8	19,3	640
35–44-v.	125	20,5	21,2	20,1	20,1	21,9	726
45–54-v.	122	20,0	19,2	22,4	22,0	23,3	772
55–64-v.	82	13,4	13,8	21,7	21,1	21,6	715
<b>Koulutus</b>							
Perusaste	129	21,1	24,5	19,5	22,5	24,8	822
Keskiaste	295	48,3	46,1	45,0	44,9	45,3	1 501
Korkea-aste	187	30,6	29,3	35,5	32,6	29,8	988
<b>Asuinlääni</b>							
Etelä-Suomi	340	55,7	60,4	38,8	42,3	41,9	1 386
Länsi-Suomi <sup>3)</sup>	271	44,4	39,6	36,3	35,4	35,4	1 171
Itä-Suomi	-	-	-	12,2	10,4	10,7	354
Oulun lääni	-	-	-	9,4	8,5	8,6	284
Lapin lääni	-	-	-	3,4	3,4	3,5	115
<b>Yhteensä</b>		<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	
<b>Yht. 1 000</b>	<b>0,611</b>	<b>0,611</b>	<b>1,031</b>	<b>4,370</b>	<b>6,745</b>		<b>3 311</b>
Etelä-S.	340	55,7	60,4	52,0	54,4	54,2	1 386
Länsi-S. <sup>3)</sup>	271	44,4	39,6	48,0	45,6	49,8	1 171
<b>Yhteensä</b>	<b>611</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>2 557</b>

<sup>1)</sup> Vastanneissa mukana pitkään (n=138), lyhyeen (n=403) ja katoaastatteluun osallistuneet (n=70).

<sup>2)</sup> 18–24-vuotiaiden osuus määriteltiin jo aikuiskoulutustutkimus 2006 otoksen poimintavaiheessa selvästi väestöosuutta pienemmäksi.

<sup>3)</sup> Länsi-Suomen läänissä on mukana Ahvenanmaa aikuiskoulutustutkimuksessa, mutta ei AKU-laaturutkimuksessa.



**Taulukko 4.6** AKU-laaturutkimuksen pitkään ja lyhyeen haastatteluun osallistuneiden jakaumat vuonna 2007 sekä aikuiskoulutustutkimukseen osallistuneet Etelä- ja Länsi-Suomen lääneissä vuonna 2006 sukupuolen, iän, koulutusasteen, alueen ja äidinkielen mukaan, %.

	AKU-LAATU			AKU 2006
	vastanneet <sup>1)</sup>	lyhyt haastattelu	pitkä haastattelu	Etelä- ja Länsi- Suomi
	(A)	(B)	(C)	(D)
<b>Sukupuoli</b>				
Mies	52,4	54,6	49,3	49,4
Nainen	47,6	45,4	50,7	50,6
<b>Ikä</b>				
18–24-v. <sup>2)</sup>	8,7	7,4	11,6	5,8
25–34-v.	37,5	38,7	34,8	30,6
35–44-v.	20,5	21,1	21,7	20,1
45–54-v.	20,0	19,9	17,4	21,1
55–64-v.	13,4	12,9	14,5	21,4
<b>Koulutus</b>				
Perusaste	21,1	18,9	25,4	19,4
Keskiaste	48,3	48,9	47,8	44,2
Korkea-aste	30,6	32,3	26,8	36,5
<b>Asuinlääni</b>				
Etelä-Suomi	55,7	57,1	55,1	52,0
Länsi-Suomi <sup>3)</sup>	44,4	42,9	44,9	48,0
<b>Äidinkieli</b>				
Suomi	89,0	89,6	87,0	92,6
Ruotsi	8,8	8,2	10,1	5,6
Muu	2,1	2,2	2,9	1,7
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Yhteensä (n)</b>	<b>611<sup>1)</sup></b>	<b>403</b>	<b>138</b>	<b>3 260</b>

<sup>1)</sup> Vastanneissa (sarake A) mukana pitkään (n=138), lyhyeen (n=403) ja katohaastatteluun osallistuneet (n=70).

<sup>2)</sup> 18–24-vuotiaiden osuus määriteltiin jo aikuiskoulutustutkimus 2006 otoksen poimintavaiheessa selvästi väestöosuutta pienemmäksi.

<sup>3)</sup> Länsi-Suomen läänissä on mukana Ahvenanmaa aikuiskoulutustutkimuksessa, mutta ei AKU-laaturutkimuksessa.

Laaturutkimuksen pitkään tai lyhyeen haastatteluun osallistuneiden jakaumat keskeisten taustamuuttujien suhteen eivät poikkea tilastollisesti merkitsevästi toisistaan. Taulukossa 4.6 on esitetty myös AKU 2006:een vastanneet Etelä- ja Länsi-Suomen lääneissä.

On tärkeää tutkia vastauskadon jakaumaa eri väestöryhmissä, mutta on kuitenkin muistettava, ettei se vielä yksistään kerro meille survey-tutkimuksella saatujen muuttujien toimivuudesta. Mielenkiinnon kohteena olevien muuttujien tulee korreloida väestömuuttujien kanssa, jotta vastauskato ja sen jakauma ennustaisi tutkittujen muuttujien hyvyttä (Peytcheva & Groves, 2009). Laaturutkimuksessa

vastauskadon jakauma on riippuvainen myös AKU 2006:n vastauskadosta (koska tutkimuksen otos on siitä poimittu).

#### 4.4 Laatututkimukseen osallistuneilla parempi työmarkkina-asema

Seuraavassa tarkastellaan, keitä ne ovat, jotka ovat osallistuneet laatututkimukseen – ja siis eivät osallistuneet AKU 2006:een. Tarkastelun kohteena on keskeisiä työelämään liittyvät muuttujia sekä vieraiden kielten ja Internetin käyttö.

Osoittautui, että laatututkimukseen osallistuneet ovat hiukan yleisemmin palkansaajia kuin AKU 2006:een osallistuneet (taulukko 4.7). Erityisesti palkansaajat osallistuivat lyhyeen, puhelimessa tehtyyn haastatteluun. Pitkään haastatteluun ja AKU 2006:een osallistuneiden työstatukset eivät juuri poikenneet toisistaan. Kun vastaajien ikä vakioidaan, erot laatututkimuksen ja AKU 2006:n välillä häviävät (ikävakioitu palkansaajien osuus laatututkimuksessa oli 61 prosenttia).

Laatututkimukseen osallistuneet palkansaajat työskentelivät AKU 2006:een osallistuneita hiukan useammin yksityissektorilla ja harvemmin kuntasektorilla. Eroja työnantajasektorissa ei ilmennyt pitkään ja lyhyeen haastatteluun osallistuneiden palkansaajien kesken. Nämä erot säilyivät iän vakioimisen jälkeenkin.

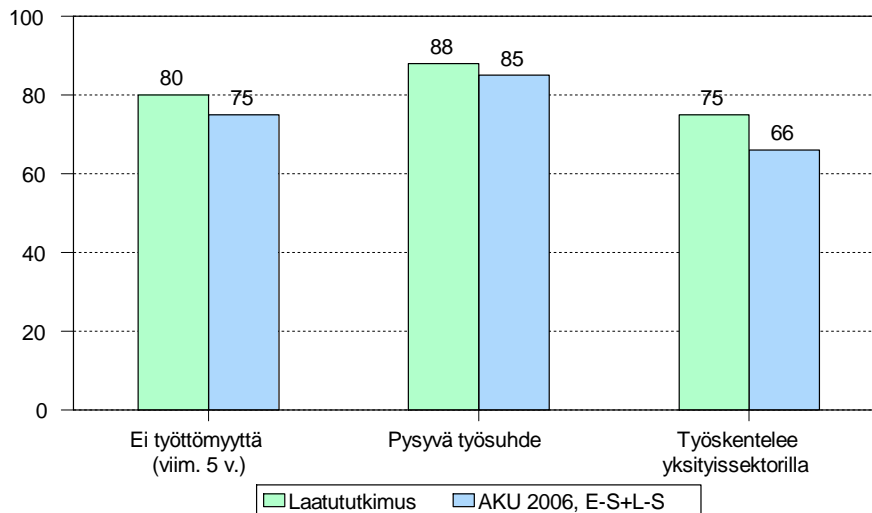
**Taulukko 4.7** Laatututkimukseen ja aikuiskoulutustutkimus 2006:een osallistuneet työstatuksen mukaan, %.

	AKU-laatututkimus 2007			AKU 2006		
	Pitkä haastattelu	Lyhyt haastattelu	Kaikki	Etelä- ja Länsi-Suomi	Kaikki	Kaikki, painotettu
Palkansaaja	66	72	71	64	61	59
Yrittäjä	7	10	9	9	9	9
Työtön	4	4	4	6	7	7
Eläkkeellä	8	5	6	9	10	10
Muu	15	9	10	12	13	15
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
(n)	(138)	(403)	(541)	(3 260)	(4 370)	

<sup>1</sup> Tilastollisesti merkitseviä eroja: palkansaajia laatututkimus–AKU 2006.

Laatututkimukseen osallistuneilla oli myös aavistuksen parempi työmarkkina-asema kuin vuotta aikaisemmin varsinaiseen aikuiskoulutustutkimukseen osallistuneilla (kuvio 4.2). Ensinnäkin heillä, erityisesti lyhyeen haastatteluun osallistuneilla oli ollut vähemmän työttömyyskokemuksia viimeksi kuluneiden viiden vuoden aikana. Lisäksi laatututkimukseen osallistuneilla oli vastaamishetkellä (vuonna 2007) jonkin verran useammin pysyvä työsuhde kuin AKU 2006:een osallistuneilla. Eroista jälkimmäinen säilyi ikävakioinninkin jälkeen, mutta työttömyyskokemuksissa vakiointi iän suhteen hävitti erot laatututkimuksen ja AKU 2006:n välillä (ikävakioitu työttömyyskokemuksia kokemattomien osuus oli 74 prosenttia).

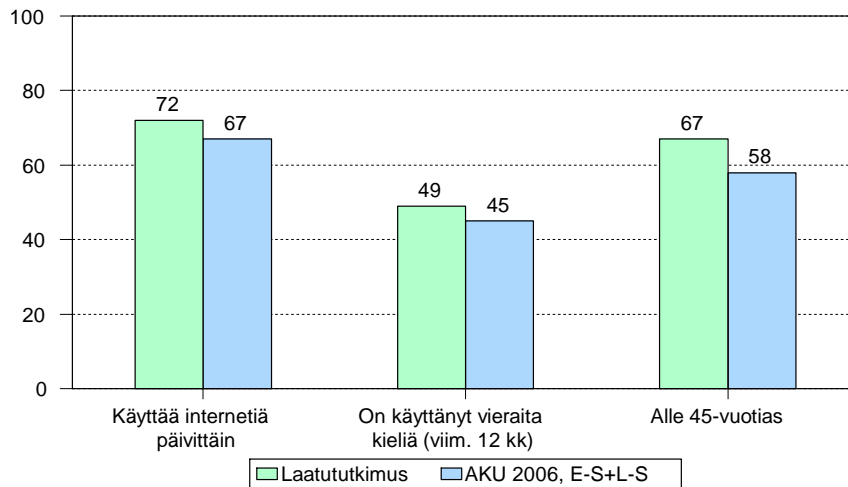
**Kuvio 4.2** Eräitä työmarkkina-asemaan liittyviä seikkoja laatututkimuksessa ja AKU 2006:ssa<sup>1</sup>, %.



<sup>1</sup> Tilastollisesti merkitseviä eroja: yksityinen sektori laatututkimus–AKU 2006 (työttömyys lyhyt–AKU 2006, pysyvä työsuhde pitkä–AKU 2006 ja lyhyt–AKU 2006).

Myös Internetin käyttö oli yleisempää laatututkimukseen osallistuneiden keskuudessa kuin vuotta aikaisemmin aikuiskoulutustutkimukseen osallistuneilla (kuvio 4.3). Tosin tämäkin yhteys katoaa, kun ikä vakioidaan (ikävakioitu päivittäin Internetiä käyttävien osuus oli 61 ja ei-käyttävien osuus 10 prosenttia). Tieto- ja viestintätekniikan käyttö -tutkimuksen mukaan Internetin käyttö on 2000-luvun jälkipuoliskolla lisääntynyt vuosittain noin 2–4 prosenttiyksikköä (<http://www.tilastokeskus.fi/til/sutivi>). Näin ollen osa erosta voinee selittyä myös yhteiskunnassa tapahtuneilla muutoksilla.

**Kuvio 4.3** Internetin käyttö (päivittäin), vieraiden kielten käyttö<sup>2,3</sup> (viimeksi kuluneiden 12 kuukauden aikana) ja alle 45-vuotiaiden osuudet laatututkimuksessa ja AKU 2006:ssa<sup>1</sup>, %.



<sup>1</sup> Tilastollisesti merkitseviä eroja: (Internet lyhyt–AKU 2006, kielet pitkä–AKU 2006).

<sup>2</sup> Vieraiden kielten käyttö töissä tai vapaa-aikana kysyttiin lyhyessä haastattelussa yhdellä kysymyksellä, mutta pitkässä haastattelussa ja AKU 2006:ssa kielten käyttöä työssä ja vapaa-aikana kysyttiin erikseen. Tulokset eivät siis ole täysin vertailukelpoisia.

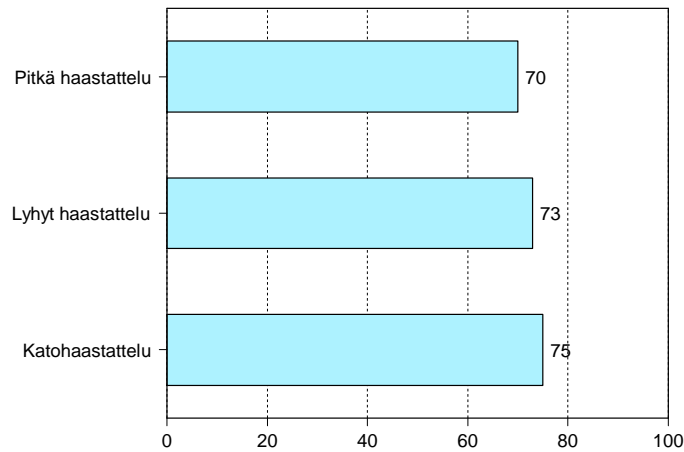
<sup>3</sup> Seitsemän pitkään haastatteluun osallistuneen vastaustiedot vieraiden kielen käyttöön puuttuvat.

#### 4.5 Haastatteluun osallistuneet äänestivät eduskuntavaaleissa

Yksi laatututkimuksen katokysymys käsitteli eduskuntavaaleissa äänestämistä maaliskuussa 2007 eli juuri ennen laatututkimuksen tiedonkeruuta. Eduskuntavaalien äänestysaktiivisuus koko maassa oli 67,9 prosenttia. Runsas 70 prosenttia laatututkimukseen osallistuneista etelä- ja länsisuomalaisista ilmoitti äänestäneensä eduskuntavaaleissa. Ilmoitettu äänestysaktiivisuus ei vaihdellut sen mukaan, mihin laatututkimuksen haastatteluun henkilö osallistui (kuvio 4.4).

Koska laatututkimuksen otos on osa aikuiskoulutustutkimus 2006:n vastauskatoa, niin tähän tutkimukseen osallistuneet ovat samalla sekä katotapauksia että tutkimukseen osallistuneita. Saatu tulos (itse ilmoitettu äänestysaktiivisuus hiukan suurempi kuin virallinen äänestysprosentti Suomessa) ei siis suoranaisesti tue ajatusta, että äänestämättä jättäminen ja tutkimuksiin osallistumattomuus olisivat ilmiöitä, jotka kasaantuisivat samoille henkilöille (esim. Tourangeau et al. 2010). Toisaalta on tiedossa, että haastattelututkimuksilla on taipumusta yliestimoida äänestyskäyttäytymistä. Tämä perustuu oletuksiin, että ei-äänestäjät jäävät helposti katoon ja/tai sosiaalisilla paineilla ”valehdella” haastatteluissa äänestäneensä (esim. Tourangeau et al. 2010, Voogt & Saris 2005, Grandberg & Homberg 1991).

**Kuvio 4.4** Eduskuntavaaleissa maaliskuussa 2007 äänestäneiden osuus tiedonkeruumuodon mukaan, %. (Mukana vain ne, jotka vastasivat kysymykseen<sup>1</sup> ja joilla oli tuolloin äänioikeus.)



<sup>1</sup> Haastattelija sai itse tapauskohtaisesti arvioida, oliko kysymyksen esittäminen mahdollista ns. katohaastattelussa.

#### 4.6 Tutkimukseen osallistuminen logit-mallituksen valossa

Tutkimukseen osallistumista ja katoon jäämistä tarkasteltiin myös logistisen regressioanalyysin, erityisesti logit-mallituksen avulla. Tällöin tutkimuskysymys 1 (sivu 22) jaettiin kahdeksi analyysiksi:

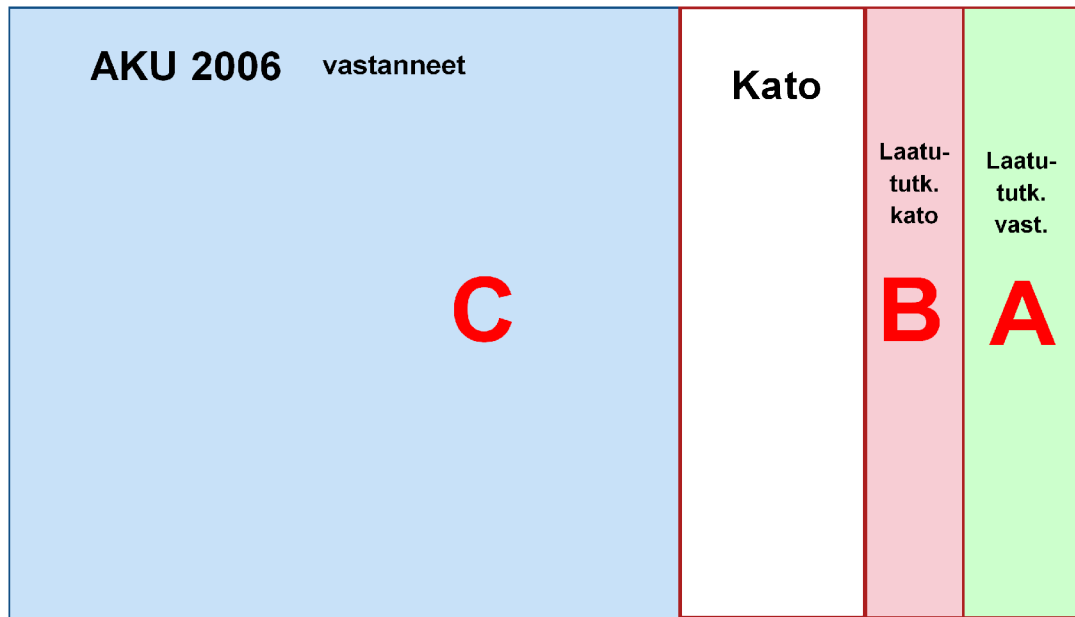
(1) Ensinnäkin verrattiin laatututkimukseen vastanneita ja sen vastauskatoa. Tätä kuvataan kuviossa 4.5 lohkojen A ja B vertailuna. Analyysimuuttujina on mahdollista käyttää rekisteritietojen lisäksi tietoja vuoden 2006 osallistumisesta (vastauskadon syyt, haastattelija, yhteydenoton ajankohta yms.)

*Analyysi 1:  $A \sim B$ , muuttujat: rekisteritiedot + tietoja vuoden 2006 osallistumisesta*

(2) Toiseksi verrattiin AKU 2006:n vastanneita laatututkimukseen vastanneisiin. Tällöin laatututkimukseen vastanneet nähdään vastauskadon edustajina. Etuna tässä ajattelussa on se, että pystymme hyödyntämään rekisteritietojen lisäksi AKU 2006:n (lohko C) ja laatututkimuksen (osa lohkoista A) vastaustietoja. Tässä analyysissä (toisin kuin muulla tässä tutkimusraportissa) huomioidaan laatututkimuksen vastanneista vain pitkään tai lyhyeen haastatteluun osallistuneet (n=541), koska katohaastattelun osalta (n=70) ei tarvittavia vastaustietoja ole saatavissa.

*Analyysi 2:  $C \sim A_{pitkä, lyhyt}$ , muuttujat: rekisteritiedot + vastaustiedot*

**Kuvio 4.5** AKU 2006:n aineisto (C, sin) ja vastauskato (valk + pun + vihreä) sekä AKU-laatatutkimuksen aineisto (A, vihreä) ja vastauskato (B, pun), Etelä- ja Länsi-Suomen läänit.



Laatatutkimukseen osallistumista (analyysi 1) selittävässä logit-mallituksessa tilastollisesti merkitseviksi selittäjiksi tulivat henkilön koulutus, asuinlääni sekä katoon jäämisen syy vuonna 2006.

Taulukossa 4.8 esitetyt tulokset on tulkittavissa siten, vain perusasteen koulutuksen omaavien riski kuulua laatututkimuksen vastauskatoon<sup>11</sup> oli 1,4-kertainen verrattuna keski- tai korkea-asteen koulutuksen omaavien riskiin. Vuonna 2006 tavoittamattomaksi jääneiden katoriski oli 1,8-kertainen verrattuna tuolloin kieltäytyneiden riskiin. Länsisuomalaisilla vetosuhde kuulua vastauskatoon oli pienempi kuin Etelä-Suomessa asuvilla (OR=0,7) tai sama asian toisin päin: eteläsuomalaisten katoriski oli 1,5-kertainen<sup>12</sup> verrattuna länsisuomalaisten riskiin jäädä laatututkimuksen katoon.

Logit-analyysissa käytetyt muuttujat, niiden luokitukset sekä logit-mallituksen täysimalli on kuvattu liitteessä 5. Logit-mallituksen etuna pelkkään frekvenssianalyysiin on se, että kun muuttuja tulee mukaan malliin, niin se vaikutus tuloksiin samalla tavallaan vakioituu.

<sup>11</sup> Selitettävänä muuttujana on laatututkimukseen osallistuminen (laatuAKU), joka saa arvot 1=osallistuu ja 0=vastauskato. SAS logistic-proseduuri mallittaa todennäköisyydet aina pienemmälle luokalle (laatuAKU=0) eli OR:t liittyvät tässä tapauksessa katoon kuulumiseen.

<sup>12</sup>  $OR = 1 / 0,671 = 1,5$ . OR:n kanteisyys perustuu kavaan 3.21.

**Taulukko 4.8** Laatututkimukseen osallistuminen versus vastauskatoon kuuluminen, logit-mallitus<sup>1</sup> ( $A \sim B$ ). Tilastollisesti merkitsevät muuttujat ja niihin liittyvät regressiokertoimet, p-arvot ja vedonlyöntisuhteet (OR).

Muuttuja	Estimaatti	Standard Error	Pr > ChiSq	OR
<b>vakio</b>	-0,5922	0,1143	0,0062	
<b>koulu1</b>	0,3470	0,1491	0,0200	1,415
<b>loptulos06</b>	0,5781	0,1300	<0,0001	1,783
<b>laani</b>	-0,3990	0,1350	0,0031	0,671

<sup>1</sup> *Malli: laatuAKU= vakio + koulu1 + loptulos06 + laani.*

Selitettävänä muuttujana laatututkimukseen osallistuminen laatuAKU (1=osallistuu, 0=vastauskato). SAS logistic mallittaa arvot luokalle 0 eli vastauskatoon kuulumisen todennäköisyyksiä.

Havaintoja n = 1 031

Kun tutkimukseen osallistumista ja katoon kuulumista mallitettiin rekisteritietojen lisäksi myös saatujen vastaustietojä hyödyntäen, käytettiin selitettävänä muuttujä AKU 2006:een osallistumista (analyysi 2). Tällöin laatututkimus tulkittiin vastauskadoksi ja siitä voitiin hyödyntää vain pitkään ja lyhyeen haastatteluun osallistuneiden tietojä.

Malliin saatiin selittäjiksi vastaajan koulutustaustan, iän ja äidinkielen lisäksi työsuhteen pysyvyys, yksityissektorilla työskentely ja koettu työhön liittyvä lisäkoulutustarve (taulukko 4.9).

Kaikki saadut odds ratiot olivat suuruudeltaan 1,4–1,6 (kun huomioidaan dummy-muuttujien käänteisyys<sup>13</sup>). Muuta kuin suomea äidinkielenään puhuvien ”riski” kuulua AKU 2006:n vastauskatoon oli 1,6-kertainen verrattuna suomea äidinkielenään puhuvien riskiin. Yksityissektorilla työskentelevien (A26) riski kuulua tutkimuksen katoon oli 1,4-kertainen verrattuna valtiolla tai kunnan palveluksessa työskentelevien riskiin. Jos vastaajalla oli pysyvä työsuhte, oli riski kuulua vastauskatoon 1,5-kertainen määräaikaisen työsuhteen omaaviin verrattuna. Henkilöt, jotka kokivat tarvitsevansa lisää työhön liittyvää koulutusta, jäivät AKU 2006:n vastauskatoon muita todennäköisemmin (OR=1,5).

AKU 2006:een osallistuttiin hiukan muita todennäköisemmin, jos vastaaja oli yli 44-vuotias tai omasi korkea-asteen koulutuksen.

<sup>13</sup> koulu5: 1/0,662=1,51, ika3: 1/0,638=1,57 ja kieli: 1/608=1,64

**Taulukko 4.9** AKU 2006:n katoanalyysi laatututkimukseen osallistuneiden osalta eli AKU 2006:n vastanneet versus laatututkimukseen vastanneet, logit-mallitus<sup>1</sup> ( $C \sim A_{\text{pitkä, lyhyt}}$ ). Tilastollisesti merkitsevät muuttujat ja niihin liittyvät regressiokertoimet, p-arvot ja vedonlyöntisuhteet (OR).

Muut- tuja	selitys	Estimaatti	Standard Error	Pr > ChiSq	OR
<b>vakio</b>		-1,6555	0,2557	< 0,0001	
<b>koulu5</b>	kork.koulu	-0,4129	0,1210	0,0006	0,662
<b>ika3</b>	yli 44-v.	-0,4492	0,1261	0,0004	0,638
<b>kieli</b>	äidinkieli	-0,4979	0,1893	0,0086	0,608
<b>A5</b>	pysyvä työsuhde	0,3854	0,1797	0,0320	1,470
<b>A26</b>	yksityissektori	0,3027	0,1336	0,0232	1,353
<b>J1</b>	koulutustarve	0,3918	0,1158	0,0007	1,480

<sup>1</sup>Malli:  $aku06 = vakio + koulu5 + ika3 + kieli + A5 + A26 + J1$

Selitettävänä muuttujana AKU 2006:een osallistuminen  $aku06$  (1=osallistuu, 0=vastauskato eli laatututkimuksen pitkään tai lyhyeen haastatteluun osallistuneet).

Havaintoja  $n = 2\,451$

Koska osa edellä kuvatun logit-mallituksen selittäjistä koskee vain työelämässä mukana olevia (esimerkiksi työsuhteen pysyvyys) ja koska SAS:in proc logistic -ohjelma ottaa logit-mallitukseen mukaan vain ne havainnot, joissa ei ole puuttuvia tietoja, oli aineiston koko edellä esitetyssä analyysissä vain 64 prosenttia koko aineistosta ( $n = 2\,451$ ).

Kun vastaava mallitus tehtiin vain niiden muuttujien avulla, joissa ei ollut puuttuvia tietoja eli käytännössä kaikki työsuhteeseen liittyvät vastausaineiston muuttujat pudotettiin pois mallista, tuli selittäjiksi vastaajan koulutustaustan, iän ja äidinkielen lisäksi asuinläänin sekä vastaustiedoista ammattiasema (A1) ja harrastuksiin liittyvä lisäkoulutustarve (J3, taulukko 4.10).

Edelleen korkea-asteen koulutuksen omaavilla, suomea äidinkielenään puhuvilla ja 45 vuotta täyttäneillä oli muita pienempi todennäköisyys jäädä tutkimuksen katoon. Tämän lisäksi vastauskatoon jäämisen riski oli pienempi Länsi-Suomen kuin Etelä-Suomen läänissä asuvilla.

Palkansaaajien ja yrittäjien riski joutua vastauskatoon oli 1,6-kertainen verrattuna työttömien ja työelämän ulkopuolella olevien katoriskiin. Sen sijaan ne, jotka kokivat



harrastuksiin liittyvää koulutustarvetta, joutuivat vastaukseen harvemmin (OR=0,8) kuin lisäkoulutustarvetta kokemattomat.

Tulokset vahvistavat frekvenssianalyysissa esiin nousseen tulokinnan siitä, ettei vastaukseen jääneillä välttämättä ole tutkimukseen osallistujia huonompi yhteiskunnallinen asema (kuvio 4.2, sivu 45). Vastaukseen jäämisen riski oli suurempi yksityissektorilla työskentelevien, pysyvän työsuhteen omaavilla ja palkansaajilla ja yrittäjillä.

**Taulukko 4.10** AKU 2006:n katoanalyysi laatututkimukseen osallistuneiden osalta eli AKU 2006:n vastanneet versus laatututkimukseen vastanneet, logit-mallitus<sup>1</sup> ( $C \sim A_{\text{pitkä, lyhyt}}$ ). Mukana vain ne muuttujat, joissa ei ole puuttuvia tietoja. Tilastollisesti merkitsevät muuttujat ja niihin liittyvät regressiokertoimet, p-arvot ja vedonlyöntisuhteet (OR).

Muut- tuja	selitys	Estimaatti	Standard Error	Pr ChiSq	OR
<b>vakio</b>		-1,1126	0,2246	< 0,0001	
<b>koulu5</b>	kork.koulu	-0,3474	0,1028	0,0007	0,706
<b>laani</b>	Länsi-Suomi	-0,1880	0,0947	0,0471	0,829
<b>ika3</b>	yli 44-v.	-0,4390	0,0994	<0,0001	0,645
<b>A1</b>	palkansaaja	0,4439	0,1166	0,0001	1,559
<b>J3</b>	koulutustarve	-0,2080	0,1013	0,0400	0,812
<b>kieli</b>	suomi	-0,4349	0,1542	0,0048	0,647

<sup>1</sup>Malli:  $aku06 = vakio + koulu5 + ika3 + A1 + J3$

Selitettyinä muuttujana AKU 2006:een osallistuminen aku06 (1=osallistuu, 0=vastaukseen eli laatututkimukseen pitkään tai lyhyeen haastatteluun osallistuneet).

Havaintoja n = 3 793

## *B Katokysymykset*

Kaikilta johonkin laatututkimuksen haastatteluun (pitkä, lyhyt tai katohaastattelu) osallistuneilta kysyttiin kymmenkunta (9–15 kappaletta) Tilastokeskusta ja tutkimukseen osallistumista koskevaa kysymystä, niin sanottua katokysymystä. Katokysymykset on esitetty liitteessä 1.

Osa kysymyksistä on samoja tai lähes samoja kuin vuonna 1989 tehdyssä tulonjakotutkimuksen palautekyselyssä. Tuolloin lähetettiin joka kymmenennelle (569) samana vuonna tulonjakotutkimukseen osallistuneelle kotitaloudelle postikyselylomake. Lomakkeita palautui 391 kappaletta (68 prosenttia). (Virtanen, 1991.) Vaikka tiedot on kerätty eri aikoina, eri tiedonkeruumenetelmin ja erilaisessa haastatteluasetelmassa, niin on mielenkiintoista katsoa vuoden 1989 tuloksia AKU-laatututkimuksen tuloksien yhteydessä. Tuloksien tulkinnassa tulee kuitenkin olla maltillinen.

### *4.7 Tilastokeskus tunnetaan ja arvostusta löytyy*

Laatututkimuksessa kysyttiin kaikilta tutkimukseen osallistuneilta, tunsivatko he Tilastokeskuksen entuudestaan. Useampi kuin neljä vastaajaa viidestä ilmoitti tunteneensa Tilastokeskuksen ennen vuoden 2006 yhteydenottoa. Parhaiten Tilastokeskuksen tunsivat lyhyeen tai katohaastatteluun osallistuneet – siis henkilöt, jotka valitsivat vuonna 2007 ”helpoimman” tien.

Sen sijaan ”raskaan tien” eli pitkään, käyntihaastatteluun osallistuneet pitivät muita useammin Tilastokeskusta tarpeellisena koko yhteiskunnan kannalta. Ilahduttavaa on, että tarpeettomana työnantajaamme ja työtämme piti vain murto-osa vastaajista.

**Taulukko 4.11** Tilastokeskuksen tuttuus tiedonkeruumuodon mukaan, %.

	Pitkä haast.	Lyhyt haast.	Katohaast.	Kaikki vastanneet
TK entuudestaan tuttu	77	84	85	82
TK ei entuudestaan tuttu	23	15	15	17
Ei osaa sanoa	0	1	0	1
<b><i>Yhteensä</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>
(n)	(138)	(403)	(70)	(611)

<sup>1</sup> Erot tilastollisesti merkitseviä: \*.

**Taulukko 4.12** Tilastokeskuksen tarpeellisuus tiedonkeruumuodon mukaan, %.

Tilastokeskus on...	Pitkä haast.	Lyhyt haast.	Katohaast.	Kaikki vastanneet
tarpeellinen koko yhteiskunnan kannalta	80	70	61	71
hyödyksi vain pienelle osalle	16	22	27	21
tarpeeton laitos	2	2	6	3
ei osaa sanoa	2	5	6	5
<b><i>Yhteensä</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>
<b>(n)</b>	<b>(138)</b>	<b>(403)</b>	<b>(70)</b>	<b>(611)</b>

<sup>1</sup> Erot tilastollisesti merkitseviä: \*.

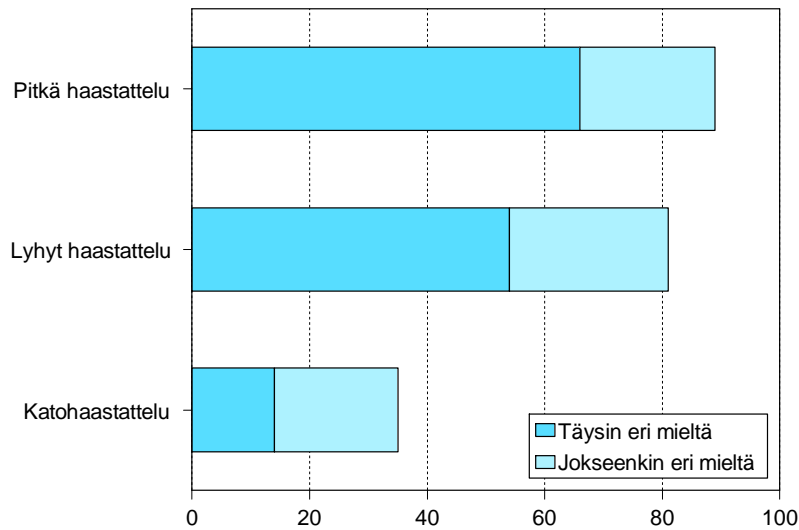
#### 4.8 Tutkimuksia pidetään tärkeinä

Kansalaisten mielipiteet tutkimuksiin osallistumisen rasittavuudesta, tutkimusten tärkeydestä ja kysymysten mahdollisesta liiasta henkilökohtaisuudesta vaihtelivat selkeästi sen mukaan, osallistuiko henkilö tutkimukseen (pitkä tai lyhyt haastattelu) vai vain katohaastatteluun.

Suuri osa laatututkimukseen osallistuneista ei koe Tilastokeskuksen haastatteluja liian tungettelevaksi. Selvä enemmistö, noin 81–89 prosenttia tutkimuksen sisältöön tutustuneista (eli pitkään tai lyhyeen haastatteluun osallistuneista) ei pitänyt kysymyksiä liian henkilökohtaisina. Sen sijaan vain katohaastatteluun suostuneista – eli henkilöistä, joilla ei ollut varsinaisesti kokemusta, vain mielikuva tutkimuksen sisällöstä – erimielisiä liian henkilökohtaisten asioiden kysymisen kanssa oli enää noin joka kolmas (kuvio 4.5). Vuonna 1989 joka kolmas vastaaja arveli vähintäänkin joka toisen suomalaisen pitävän tutkimuksia liian henkilökohtaisina<sup>14</sup>. Kolmannes ei osannut tuolloin vastata ja kolmannes arveli harvemman kuin joka toisen suomalaisen kokevan, että tutkimuksissa kysytään liian henkilökohtaisia asioita.

<sup>14</sup> Vuoden 1989 tulonjakotutkimuksen palautekyselyssä väitteet oli esitetty erilaisessa muodossa kuin laatututkimuksessa: ”Kuinka suuri osuus suomalaisista on mielestänne väitteen kanssa samaa mieltä?” Vastausvaihtoehdot olivat: selvä enemmistö, yli puolet, vaikea sanoa, alle puolet ja selvä vähemmistö. Muilta osin käytettiin laatututkimuksessa samoja väitteitä kuin vuoden 1989 tutkimuksessa. Tulonjakotutkimuksen palautekyselyn tulosten vertailtavuus nykypäivään on vähintäänkin ongelmallista – auki kirjoittamisesta puhumattakaan.

**Kuvio 4.5** Väitteen "*Tutkimuksissa kysytään liian henkilökohtaisia asioita*" kanssa täysin tai jokseenkin eri mieltä olevien osuudet tiedonkeruumuodon<sup>1</sup> mukaan, %.

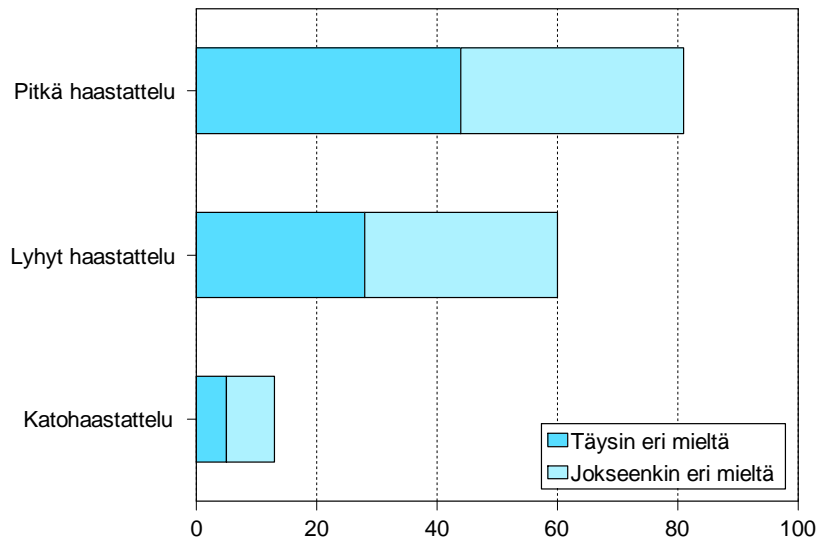


<sup>1</sup> Tässä katohaastattelun osalta tarkasteluun on otettu mukaan vain ne, jotka ovat vastanneet kysymykseen.

Myös näkemys siitä, viekö tutkimuksiin osallistuminen liikaa aikaa vai ei, poikkesi selkeästi sen suhteen, osallistuiko henkilö vain katohaastatteluun, lyhyeen haastatteluun vai pitkään haastatteluun (kuvio 4.6). Katohaastatteluun osallistuneista jopa 73 prosenttia piti osallistumista ajallisesti rasittavana. Lyhyeen haastattelun osallistuneista tätä mieltä oli noin joka neljäs ja eniten aikaansa uhranneista eli pitkään haastatteluun osallistuneista vain 7 prosenttia.

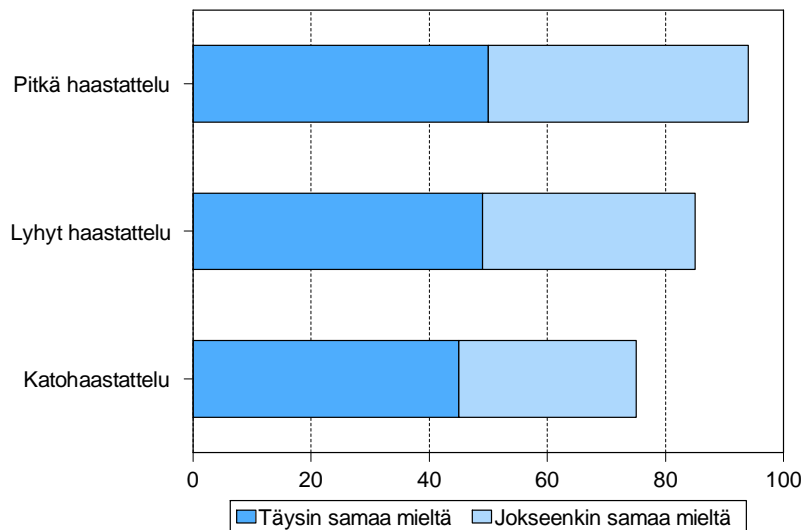
Selvä enemmistö 18–64-vuotiaista laatututkimukseen osallistuneista näki, että tutkimuksilla saadaan arvokasta tietoa suunnittelua varten (kuvio 4.7). Tämän väitteen osalta samaa mieltä olevien osuus oli hiukan suurempi pitkään tai lyhyeen haastatteluun osallistuneiden kuin katohaastattelun antaneiden joukossa, mutta erot eivät olleet enää niin silmiinpistäviä. Vuonna 1989 joka toinen tutkimukseen osallistunut arveli, että joka toinen suomalainen arvostaa tutkimuksella saatua tietoa<sup>7</sup>.

**Kuvio 4.6** Väitteen "*Tutkimuksiin osallistuminen vie liian paljon aikaa*" kanssa täysin tai jokseenkin eri mieltä olevien osuudet tiedonkeruumuodon<sup>1</sup> mukaan, %.



<sup>1</sup> Tässä katohaastattelun osalta tarkasteluun on otettu mukaan vain ne, jotka ovat vastanneet kysymykseen.

**Kuvio 4.7** Väitteen "*Tutkimuksilla saadaan arvokasta tietoa suunnittelua varten*" kanssa täysin tai jokseenkin samaa mieltä olevien osuudet tiedonkeruumuodon<sup>1</sup> mukaan, %.



<sup>1</sup> Tässä katohaastattelun osalta tarkasteluun on otettu mukaan vain ne, jotka ovat vastanneet kysymykseen.

Pitkään haastatteluun osallistuneilta kysyttiin myös mielipidettä väitteistä:

- (A) "*Yksityisen kansalaisen antamilla tiedoilla ei ole mitään merkitystä*",
- (B) "*Vastaamalla tutkimukseen yksityinen henkilö voi vaikuttaa asioihin*" ja
- (C) "*Tutkimuksiin osallistuminen antaa ajattelemisen aihetta omankin elämän kannalta*".

Lisäksi heitä pyydettiin arvioimaan tilastotiedon ymmärrettävyyttä:

(D) ”*Tilastotiedon ymmärrettävyys tavalliselle kansalaiselle on helppoa, melko helppoa, melko vaikeata vai lähes mahdotonta.*”

Selvä enemmistö (81 prosenttia) piti kansalaisten antamia tietoja merkityksellisinä (eli oli eri mieltä ensimmäisen väitteen kanssa). Kolme neljästä (76 prosenttia) uskoi yksityisen henkilön vaikutusmahdollisuuksiin ja näki, että osallistuminen antaa ajattelemisen aihetta myös oman elämän kannalta. (Taulukko 4.13.)

Lisäksi neljä vastaajaa viidestä (pitkä haastattelu) piti tilastotiedon ymmärrettävyyttä vähintäänkin melko helppona. Huolestuttavaa kuitenkin on, että noin joka viides vastaaja näki tavallisen kansalaisen olevan kovilla tuottamiemme tietojen ymmärtämisen suhteen.

**Taulukko 4.13** Eräiden tutkimuksen osallistumista koskevien väitteiden vastausjakaumat, %. Väitteet on esitetty vain pitkässä haastattelussa.

	Kansalaisen antamalla tiedoilla ei merkitystä	Vastaamalla voi vaikuttaa asioihin	Osallistuminen antaa ajattelemisen aihetta
	(A)	(B)	(C)
Täysin samaa mieltä	4	37	36
Jokseenkin samaa mieltä	7	39	40
Ei samaa, ei eri mieltä	9	15	12
Jokseenkin eri mieltä	33	5	9
Täysin eri mieltä	48	3	1
<b>Yhteensä</b>	<b>100</b>	<b>100</b>	<b>100</b>
(n)	(138)	(138)	(138)

#### 4.9 Päätös osallistua syntyi kirjeen ja haastattelijan soiton myötä

Joka toinen tutkimukseen osallistunut kertoi tehneensä osallistumispäätöksensä vasta haastattelijan soiton yhteydessä ja joka kolmas päätti osallistua jo saatuaan saatekirjeen. Vain 13 prosenttia jäi odottamaan lisätietoa ennen osallistumispäätöksen tekoa. Tulos kertoo yhtäältä haastattelijoiden kontaktin tärkeydestä ja onnistumisesta, mutta myös saatekirjeiden merkityksestä (ja toimivuudesta). Osallistumispäätöstä koskevat kysymykset esitettiin vain pitkässä eli käyntihaastattelussa. Tulokset eivät poikkea 1980-luvun lopulla saaduista vastauksista.

Vastaajilta kysyttiin lisäksi kolmea syytä, jotka vaikuttivat osallistumispäätökseen. Yleisin syy oli käsitys siitä, että (a) tutkimuksella saadaan tärkeitä tietoja. Joka neljäs osallistumissyö oli tätä. Myös (b) asioihin vaikuttaminen ja tutkimuksen kiinnostavuus kuuluivat yleisimpien osallistumissyiden joukkoon. Tutkimusta koskevien syiden lisäksi osallistumispäätökseen vaikuttivat se, että (c) vastaajalla sattui olemaan aikaa (15 prosenttia syistä), (d) haastattelijan ammattitaitoisuus (joka kymmenes syy) sekä se, (e) ettei haastateltava tohtinut kieltäytyä (7 prosenttia syistä). (Taulukko 4.14.)

**Taulukko 4.14** Osallistumispäätökseen vaikuttavia seikkoja (pitkä haastattelu), prosenttiosuudet annetuista vastauksista (A) ja vastanneista (B). Jokainen vastaaja sai nimetä kolme seikkaa.

	(A) % vastauksista	(B) % vastanneista
Tutkimuksella saadaan tärkeitä tietoja	23	62
Sattui olemaan vapaata aikaa	15	40
Tutkimus vaikutti kiinnostavalta	14	38
Voi vaikuttaa asioihin	10	28
Haastattelija vaikutti ammattitaitoiselta	10	25
Ei tohtinut kieltäytyä	7	19
On hauska keskustella ihmisten kanssa	6	15
Tutkimuksiin vastaaminen on jokaisen velvollisuus	5	12
Tutkimuksen teki valtion laitos	4	11
Vastaa yleensä kaikkiin tutkimuksiin	3	7
Perheenjäsenet, ystävät tai tuttavat kehottivat	1	3
Jokin muu asia	2	7
<b><i>Yhteensä</i></b>	<b><i>100</i></b>	
Vastauksia yhteensä	(367)	
Vastaajia yhteensä		(138)

Tilastokeskuksessa vastaajia on perinteisesti pyritty motivoimaan sillä, että tutkimuksen tekee Tilastokeskus, valtion laitos. Osallistuminen tutkimuksiin tulisi olla siten jollakin tavoin lähes velvollisuus. Molemmat perustelut (Tilastokeskus, velvollisuus) edustivat vain noin viittä prosenttia vastaajien ilmoittamista osallistumissyistä. Toisin oli vuonna 1989, jolloin valtion laitos tutkimuksen tekijänä oli yksi viidestä tärkeimmästä osallistumispäätökseen vaikuttavasta seikasta. Muilta osin viiden kärki osallistumissyissä oli samanlainen vuonna 1989 ja lähes kaksikymmentä vuotta myöhemmin vuonna 2007.

Laatututkimuksessa osallistumissyitä koskeva kysymys oli lomakkeen lopussa ja vastausvaihtoehdot esitettiin vastaajille vastauskorteilla. Jokainen sai nimetä kolme tärkeintä osallistumispäätökseensä vaikuttanutta seikkaa.

#### 4.10 Aikaisempi yhteydenotto muistettiin

Haastattelijoita pyydettiin arvioimaan jokaisen haastattelun lopuksi, muistiko haastateltava aikaisemman vuonna 2006 tapahtuneen yhteydenoton aikuiskoulutustutkimuksen osalta. Noin joka toinen laatututkimukseen osallistunut muisti haastattelijan arvion mukaan aikaisemman yhteydenoton. Jos yhteydenotto oli jo unohtunut, oli osallistuminen pitkään käyntihaastatteluun todennäköisempää kuin lyhyeen haastatteluun. Tämä ero oli tilastollisesti merkitseviä. (Taulukko 4.15.)

Haastattelijaa pyydettiin arvioimaan muistamista myös niiden osalta, jotka jäivät laatututkimuksessa(kin) vastauskatoon. Tehtävä ei ollut helppo, sillä osa kontakteista oli kovin lyhyitä (päättös kieltäytymisestä nopeasti) tai jäi kokonaan saamatta (ei-tavoitetut). Haastattelijaa arvioi kuitenkin 57 laatututkimuksessa vastauskatoon jääneen (14 prosenttia katotapauksista) osalta aikaisemman yhteydenoton muistamista. Näistä he arvioivat lähes 70 prosentin muistaneen vuoden 2006 yhteydenoton.

**Taulukko 4.15** Haastattelijan arvio siitä, muistiko haastateltava aikuiskoulutustutkimuksen yhteydenoton vuodelta 2006 tiedonkeruumuodon mukaan, %. Tässä huomioidaan vain ne tapaukset, joissa muistaminen tai muistamattomuus kävi ilmi.

<i>Muisti aikaisemman yhteydenoton...</i>	Pitkä haast.	Lyhyt haast.	Kato-haast.	Kato-tapaukset	Kaikki
Kyllä	56	65	67	68	64
Muisti myöhemmin	13	18	7	18	16
Ei ollut varma	10	8	12	9	9
Ei muistanut	22	10	14	5	12
<b><i>Yhteensä</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>	<b><i>100</i></b>
(n)	(125)	(357)	(58)	(57)	(597)

Haastattelijoiden arvio ei voinut perustua muistiin, sillä haastattelijaa-haastateltava-pari pyrittiin vaihtamaan AKU 2006:n ja laatututkimuksen välillä. Näin ollen haastateltava oli haastattelijalle vain harvoin entuudestaan tuttu: 85 prosenttia (906 kappaletta) laatututkimuksen yhteydenotoista teki eri haastattelijaa kuin vuonna 2006.



Yllättävää ei ole, että henkilöt, joita ei tavoitettu vuonna 2006 aikuiskoulutustutkimuksen yhteydessä, muistivat vuoden kuluttua ”yhteydenoton” (esimerkiksi saatekirjeen) huomattavasti huonommin kuin vuonna 2006 tutkimuksesta kieltäytyneet (taulukko 4.16). Tosin lähes joka toinen vuonna 2006 tavoittamattomaksi jäänyt, mutta vuotta myöhemmin tutkimukseen suostunut, muisti aikuiskoulutustutkimuksen yhteydenoton.

**Taulukko 4.16** Haastattelijan arvio siitä, muistiko pitkään, lyhyeen tai katohaastatteluun vuonna 2007 suostunut aikaisemman yhteydenoton sen mukaan, kieltäytyikö henkilö vai eikö häntä tavoitettu vuonna 2006. Prosenttiosuudet tapauksista, joissa muistaminen / muistamattomuus kävi ilmi.

<i>Muisti aikaisemman yhteydenoton...</i>	Kieltäytyi vuonna 2006	Ei tavoitettu vuonna 2006
Kyllä	74	46
Ei heti, myöhemmin	13	19
Ei ollut varma	7	10
Ei muistanut	5	25
<b><i>Yhteensä</i></b>	<b><i>100</i></b>	<b><i>100</i></b>
(n)	(328)	(212)

#### 4.11 Pitäisikö haastatteluun osallistumisesta palkita?

Katokysymyksissä oli kolme palkitsemiseen liittyvää kysymystä. Palkitsemisen tarpeellisuutta kysyttiin kaikilta vastanneilta. Tarpeellisimpina palkitsemista pitivät vain katohaastatteluun suostuneet, tosin heistäkin vain 14 prosenttia piti sitä erittäin tarpeellisena. Palkitsemisen merkitys laski haastattelun ”pituuden” lisääntyessä. Suurimman ”panoksen” antaneet eli pitkään haastatteluun suostuneet näkivät useimmiten palkitsemisen melko tai täysin tarpeettomana – heistä 60 prosenttia oli tätä mieltä. Lyhyeen haastatteluun osallistuneista palkkiota piti tarpeettomana 44 prosenttia. (Taulukko 4.17.)

Vuonna 1989 tulonjakotutkimuksen palautekyselyyn vastanneet pitivät palkitsemista aavistuksen harvemmin tarpeettomana kuin vastaajat 2000-luvun ensimmäisellä vuosikymmenellä.

Palkitsemista tarpeellisena tai erittäin tarpeellisena pitäneiltä pitkään ja lyhyeen haastatteluun osallistuneilta kysyttiin myös, mikä olisi sopiva palkkio tunnin pituisesta haastattelusta. Lisäksi haastatteluun osallistuneilta tiedusteltiin, sopisiko

vastaajien palkitseminen Tilastokeskuksen haastattelututkimuksiin. Mielipiteet palkkiomuodosta vaihtelivat raha- ja tavarapalkkiosta arvontaan. Samoin palkittamisen sopivuudesta valtion laitoksen imagoon oltiin sekä puolesta että vastaan. Pitkään haastatteluun osallistuneet katsoivat muita useammin, ettei vastaajien palkitseminen sopisi Tilastokeskuksen haastattelututkimuksiin (38 prosenttia oli tätä mieltä).

**Taulukko 4.17** Mielipiteet tutkimukseen osallistumisen palkittamisesta tiedonkeruumuodon mukaan, %.

<i>Palkittaminen on...</i>	Pitkä haastattelu	Lyhyt haastattelu	Kato-haastattelu	Kaikki vastanneet	Tulonjako tutkimuksen palautekysely 1989 <sup>1</sup>
erittäin tarpeellista	3	7	14	7	11
tarpeellista	17	24	24	22	21
vaikea sanoa	20	25	27	24	22
melko tarpeetonta	28	26	14	25	22
täysin tarpeetonta	32	18	14	20	16
eos, puuttuva	1	1	6	1	7
<b><i>Yhteensä</i></b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
(n)	(138)	(403)	(70)	(611)	(392)

<sup>1</sup> Lähde: Virtanen 1989.

### *C Aikuiskoulutustutkimuksen keskeiset indikaattorit*

Seuraavassa tarkastellaan, kuinka AKU-laaturutkimukseen osallistuneet osallistuivat aikuiskoulutukseen. Kysymysten osalta eri tiedonkeruumuodot poikkesivat hiukan toisistaan. Osa tutkimuskysymyksistä on esitetty vain pitkässä haastattelussa ja osassa kysymyksistä on hiukan erilainen kysymysmuotoilu pitkän ja lyhyen haastattelun kesken.

#### *4.12 Aikuiskoulutukseen osallistutaan samassa määrin*

Laaturutkimuksessa kysyttiin osallistumista eri organisaatioiden järjestämään aikuiskoulutukseen. Vastaavanlainen kysymys on AKU 2006:ssakin ja sen perusteella määritellään eri aikuiskoulutukseen osallistuminen (Pohjanpää et al. 2008, 157). Kysymys on kuitenkin muotoiltu laaturutkimuksen lyhyessä, puhelimesta tehtävää haastattelua varten hiukan erilaisilla kuin käyntihaastattelussa. Lisäksi kaikki laaturutkimukseen osallistuneet olivat saaneet saatekirjeen mukana luettelon eri aikuiskoulutusmuodoista (liite 2). Tällaista koostetta ei varsinaisessa AKU 2006:ssa käytetty. Paperikoosteessa ja puhelinhaastattelun kysymyksessä on joitakin koulutusmuotoja yhdistetty ja lisäksi yleisimmät koulutusmuodot oli koottu kysymyksen alkuun. AKU 2006:ssa eri koulutusmuotoja oli 36 ja laaturutkimuksen puhelinhaastattelussa 16 kappaletta.

Taulukkoon 4.18 on koottu osallistuminen viiteen yleisimpään koulutusorganisaation järjestämään koulutukseen. Muissa koulutusmuodoissa osallistuneita oli alle 5–6 prosenttia vastanneista eikä eroja tiedonkeruumuotojen välillä ilmennyt. Tulokset eivät summaudu sataan prosenttiin eivätkä edes aikuiskoulutukseen osallistuneiden osuuteen, koska vastaaja on saattanut osallistua useampaan koulutukseen tutkimusta edeltäneiden 12 kuukauden aikana.

Laaturutkimuksen aineiston pitkään haastatteluun osallistuneille muodostettiin neljä aikuiskoulutukseen osallistumisindikaattoria. Lyhyen haastattelun aineistolle oli mahdollista laskea vain aikuiskoulutukseen osallistumisosuus, koska koulutuksen lisäkysymyksiä ei siinä esitetty. Taulukosta 4.19 näkyy, että AKU 2006:ta vastaavaan pitkään haastatteluun osallistuneet osallistuivat aikuiskoulutuksen eri muotoihin samassa määrin kuin AKU 2006:en osallistuneet ja että myös

puhelinhaastatteluun vastanneiden aikuiskoulutukseen osallistuminen on samaa tasoa kuin AKU 2006:ssa.

**Taulukko 4.18** Eri koulutusorganisaatioiden (viisi yleisintä) järjestämään koulutukseen osallistuminen<sup>1</sup> tutkimusta edeltäneiden 12 kuukauden aikana tiedonkeruumuodon mukaan vuonna 2007 sekä AKU 2006:ssa, % vastanneista<sup>2</sup>.

	Pitkä haastattelu	Lyhyt haastattelu	AKU 2006
Työnantajan järjestämä koulutus	30	31	27
Konferenssi tai seminaari	14	16	12
Työpaikalla annettu perehdytyskoulutus <sup>3</sup>	13	17	11
Kansalais- tai työväenopisto	7	7	11
Järjestön tai yhdistyksen järjestämä koulutus ammatillisessa oppilaitoksessa	7	8	8
(n)	(138)	(403)	(4 370)

<sup>1</sup> Eri organisaatioiden järjestämään koulutukseen osallistumista kysyttiin hiukan erilalla laatututkimuksen lyhyessä haastattelussa kuin käyntihaastatteluissa. Lisäksi kaikki laatututkimukseen osallistuneet saivat koulutusmuotoja esittelevää materiaalia etukäteen. Kysytyt koulutusorganisaatiot on esitetty liitteessä 2.

<sup>2</sup> Prosentit eivät summaudu aikuiskoulutukseen osallistumisosuuteen (AKU 2006:ssa 52 prosenttia), sillä yksittäinen vastaaja on saattanut osallistua useampaan koulutukseen.

<sup>3</sup> Tilastollisesti merkitseviä eroja: perehdytyskoulutus lyhyt-AKU 2006

**Taulukko 4.19** Aikuiskoulutukseen, ammatilliseen aikuiskoulutukseen, henkilöstökoulutukseen ja harrastustavoitteiseen aikuiskoulutukseen osallistuneet<sup>1</sup> tutkimusta edeltäneiden 12 kuukauden aikana tiedonkeruumuodon mukaan vuonna 2007 sekä AKU 2006:ssa, % vastanneista<sup>2</sup>.

Osallistuminen...	Laatututkimus		AKU 2006	
	Pitkä haastattelu	Lyhyt haastattelu	Etelä- ja Länsi-Suomi	painotettu kaikki
aikuiskoulutukseen	54	52	54	52
ammatilliseen aikuiskoulutukseen	50	-	53	51
henkilöstökoulutukseen	57	-	58	57
harrastustavoitteiseen aikuiskoulutukseen	16	-	19	18
(n)	(138)	(403)	(3 260)	(4 370)

<sup>1</sup> Ammatillisessa aikuiskoulutuksessa kantalukuna työvoima, henkilöstökoulutuksessa palkansaajat ja harrastustavoitteissa koulutuksessa kaikki muut paitsi opiskelijat ja varusmiehet.

<sup>2</sup> Tilastollisesti merkitseviä eroja: ei ole

### 4.13 ...mutta aikuiskoulutustarpeita koetaan enemmän

Haastateltavilta kysyttiin, onko heillä tarvetta saada lisää ammattitaitoa kehittävää tai uraa edistävää koulutusta sekä onko tarvetta tai kiinnostusta saada vapaa-aikaan tai harrastuksiin liittyvää koulutusta. Kysymysmuotoilut olivat laatututkimuksen haastatteluissa samat kuin AKU 2006:ssa, mutta toki lyhyessä puhelinhaastattelussa kysymyksen konteksti oli täysin erilainen. Kysymys esitettiin aina haastattelun loppupuolella, joten käyntihaastatteluissa koulutusasioista oli keritty jutustella jo puolisen tuntia.

Laatututkimukseen osallistuneet kokivat tarvitsevänsä useammin työhön liittyvää lisäkoulutusta kuin vuotta aikaisemmin AKU 2006:een osallistuneet. Vastaukset eivät vuonna 2007 poikenneet tiedonkeruutavan mukaan. Sen sijaan vapaa-aikaan liittyvää koulutusta kaipaili useimmiten pitkään haastatteluun osallistuneet. Harvinaisinta tällaiset koulutushaaveet olivat puhelinhaastatteluun osallistuneiden keskuudessa. (Taulukko 4.20.) Erot koulutustarpeissa eivät muuttuneet, kun tulokset vakioitiin iän ja koulutustaustan mukaan.

**Taulukko 4.20** Tarve saada lisää työhön tai harrastuksiin liittyvää koulutusta tiedonkeruumuodon mukaan vuonna 2007 sekä AKU 2006:ssa, % vastanneista<sup>1</sup>.

	AKU-laatututkimus		AKU 2006	
	Pitkä haastattelu	Lyhyt haastattelu	Etelä- ja Länsi- Suomi	painotettu, kaikki
Lisää ammattitaitoa kehittävää tai uraa edistävää koulutusta <sup>2</sup>	53	51	44	44
Tarvetta tai kiinnostusta saada vapaa-aikaan tai harrastuksiin liittyvää koulutusta	45	27	36	34
(n)	(138)	(403)	(3 260)	(4 370)

<sup>1</sup> Tilastollisesti merkitseviä eroja: työ pitkä-AKU 2006, lyhyt-AKU 2006, laatu-AKU 2006, vapaa-aika pitkä-lyhyt, pitkä-AKU 2006, lyhyt-AKU 2006.

<sup>2</sup> Mukana tarkastelussa vain työelämässä mukana olevat ( $n_{pitkä}=126$ ,  $n_{lyhyt}=383$ ,  $n_{aku\ 2006}=3895$ ).

Tuloksille on vaikea löytää hyvää selitystä. Koulutustarpeen muuttuminen vuoden sisällä ei liene vastaus. Tiedonkeruumenetelmäänkään liittyvistä eroista ei näyttäisi olevan kyse. Vaikuttaa siis siltä, että vastauskatoon (AKU 2006:sta) jääneet kaipaavat enemmän lisäkoulutusta – niin vapaa-aikaan kuin työhönkin liittyen – kuin tutkimukseen osallistuneet.

Hypoteettinen ajatus yhdistää laatututkimuksessa saadut tulokset AKU 2006:n tuloksiin ei tuota muutoksia koulutustarvetta kuvaaviin väestösuuksiin. Harrastuksiin ja vapaa-aikaan liittyvän koulutustarve pysyy Etelä- ja Länsi-Suomen läänissä 36 prosentissa ja ammattitaitoa kehittävää lisäkoulutustakin tarvitsisi edelleen 44–45 prosenttia etelä- ja länsisuomalaisista. Ajatusleikki osoittaa, että aineiston kasvattaminen noin viidellä sadalla vastaukseen jääneellä (17 prosentilla alkuperäisestä aineistosta) – osa jopa hiukan erilailla vastanneella – ei muuta indikaattoreiden suuruutta. Jos seurataan täysin ufoa ajatusta siitä, että jospa kaikki AKU 2006:n katotapaukset olisivat osallistuneet ja vastanneet kuten laatututkimuksen vastaajat, niin saadut kokonaisindikaattorit olisivat 37 (vapaa-ajan koulutustarve) ja 47 prosenttia (ammattitaidon lisäkoulutustarve) kyseisellä tutkimusalueella. Siis eroa ei näinkään epärealistisella jossittelulla saada kovin suureksi.

#### *4.14 Osallistuminen aikuiskoulutukseen logit-mallituksen mukaan*

Aikuiskoulutustutkimuksen keskeisin indikaattori on aikuiskoulutukseen osallistuminen. Vuonna 2006 aikuiskoulutukseen oli viimeksi kuluneiden 12 kuukauden aikana osallistunut 52 prosenttia 18–64-vuotiaista, yhteensä reilut 1,7 miljoona suomalaista (Pohjanpää et al. 2008).

Logit-mallituksessa tutkimuskysymyksestä 2 (sivu 22) tehtiin kaksi analyysia<sup>15</sup> siten, että selitettävänä muuttuja oli joko (a) AKU 2006:een osallistuminen tai (b) aikuiskoulutukseen osallistuminen. Vastaavasti toinen edellä mainituista muuttujista pyrittiin pakottamaan mukaan malliin selittäjänä.

Kummassakaan logit-mallituksessa eivät nämä malliin ”pakotetut” muuttujat oikeastaan kuuluneet malliin. Niiden merkitys selittäjinä ei missään vaiheessa ollut merkittävää (eli p-arvo aina suurempi kuin 0,05, esimerkiksi taulukko 4.21 ja liitetaulukko 6, liite 5). Tulos on tulkittavissa siten, että aikuiskoulutukseen osallistuminen ei riipu lainkaan siitä, onko henkilö osallistunut AKU 2006 - tutkimukseen vai ei. Kun huomioimme laatututkimuksen vastauskadon edustajana,

<sup>15</sup> Kuviossa 4.5 (sivu 44) analyysit tarkoittavat lohkojen C ja A vertailua, siten, että laatututkimuksen osalta mukana ovat mukana vain pitkään tai lyhyeen haastatteluun vastanneet. Katohaastattelun (n=70) osalta ei aikuiskoulutukseen osallistumisindikaattori ole laskettavissa.

niin tulos on, että vastauskatoon jääneet osallistuivat aikuiskoulutukseen samassa määrin kuin tutkimukseen osallistuneet. Tulos vahvistaa jo aikaisemmin esitetyn tulkinnan (taulukko 4.19).

**Taulukko 4.21** Aikuiskoulutukseen osallistuminen, logit-mallitus<sup>1</sup> (C ~ A<sub>pitkä, lyhyt</sub>).

Muuttuja	selitys	Estimaatti	Pr > ChiSq
<b>vakio</b>		2,4827	<0,0001
<b>aku06</b>	osallistui AKU 2006	-0,0603	<b>0,2315</b>
<b>koulu5</b>	korkea-aste	-0,6676	<0,0001
<b>sukup</b>	nainen	-0,5312	<0,0001
<b>A1</b>	palkansaaja	-1,0332	<0,0001
<b>G89</b>	vieraiden kielten käyttö	-0,3121	0,0002
<b>G2</b>	internetin käyttö	-0,7293	<0,0001
<b>J3</b>	koulutustarve	-0,3782	<0,0001

<sup>1</sup>Malli:  $akuOS = vakio + aku06 + koulu5 + sukup + A1 + G89 + G2 + J3$

Selitettävänä muuttujana aikuiskoulutukseen osallistuminen akuOS (1=osallistunut aikuiskoulutukseen, 0=ei ole osallistunut aikuiskoulutukseen viimeksi kuluneiden 12 kuukauden aikana). Pakotettuna selittäjänä on aku06 (1=aku 2006, 0=vastauskato)

Havaintoja n = 3 793

## 5. Yhteenveto ja pohdinta

### 5.1 Vastaus tutkimuskysymyksiin

AKU-laaturitkimuksella haluttiin saada tietoa vastauskadosta ja sille asetettiin kaksi tavoitetta. Haluttiin tietää, (1) keitä ne ovat, joita emme tutkimuksilla saa kiinni, ja (2) vastaavatko tutkimuksiin osallistumattomat samalla tapaa kuin osallistuneet.

Molempiin tutkimuskysymyksiin saatiin vastauksia. Saadut tulokset tukevat aikaisempaa teoriaa niin survey-tutkimukseen osallistumisesta kuin survey-tutkimuksen prosessin vaikutuksesta kansalaisten osallistumispäätöksiin. Myös ajatuksia lisätutkimuksille syntyi.

#### *Ei totaalikieltäytyjiä*

Ensinnäkin laaturitkimus osoitti, etteivät vastauskatoon kuuluneet ole välttämättä tutkimusmaailman ”totaalikieltäytyjiä”. Tutkimuksen perusteella voidaan todeta, että jo kertaalleen tutkimuksesta suorasti tai epäsuorasti kieltäytyneitä voidaan houkutella uudestaan mukaan hyvinkin onnistuneesti. Tästä kertoo erityisen hyvä vastausaktiivisuus: 59 prosenttia suostui yhteistyöhön ja antoi tietoa vähintäänkin katohaastattelun verran. Laaturitkimuksessa osallistumista kasvatti uuden yhteydenoton lisäksi epäilemättä myös mahdollisuus valita käyntihaastattelun sijaan lyhyempi puhelinhaastattelu. Puhelinhaastattelun valitsikin kaksi kolmesta tutkimukseen osallistuneesta.

Vastaavia tuloksia osallistumisesta on saatu muuallakin. Esimerkiksi Voogt’n ja Sarisin (2005) tutkimuksessa jo kertaalleen kieltäytyneistä osallistui tutkimukseen uuden yhteydenoton jälkeen 20–91 prosenttia.

Haastattelijoilta saadun palautteen mukaan ainakin osa laaturitkimukseen osallistunut suostui yhteistyöhön tajuttuaan, että oma vastaaminen on ”*niin tärkeitä*” [että vaivataan uudestaan]. Tällainen kokemus yhdessä lähes 60 prosentin vastausosuuden kanssa herättää ajatuksen siitä, olemmeko saatekirjeen ja haastattelijan motivointisoiton kanssa hiukan myöhässä. Tieto ”tiedontuottamistalkoista” ja vastaajan tärkeydestä tulisi olla kansalaisilla jo paljon aikaisemmin kuin posti tuo tiedon otokseen pääsemisestä.



Uusintakierroksella leppyivät ja tutkimukseen suostuivat hiukan useammin länsisuomalaiset ja vähintään keskiasteen koulutuksen saaneet. Sen sijaan ikä tai sukupuoli ei vaikuttanut osallistumisaktiivisuuteen.

Vuoden 2006 vastauskatoon jäämisen syy oli selkein osallistumista ennakoiva seikka: kieltäytyneistä 67 prosenttia, mutta tavoittamatta jääneistä vain 51 prosenttia antoi jotain tietoja laatututkimuksessa vuonna 2007. Logit-mallituksen ”kielellä” tulos oli, että tavoittamattomaksi jääneiden katoriski oli 1,8-kertainen verrattuna vuonna 2006 kieltäytyneiden riskiin. Yhdessä sen seikan kanssa, että juuri tavoittamattomien osuudet ovat kasvussa, antaa tulos aihetta pohtia katotyön suuntaamisesta survey-tutkimuksissa.

### *...vaikka osalla tutkimuksien välttäminen pysyvää*

Toisaalta laatututkimus kertoi myös vastauskadon pysyvyydestä. Lopulliseen vastauskatoon jäämisen syyt olivat usein samat kuin aikaisemmalla tiedustelukerralla – kieltäytyneet kieltäytyivät ja tavoittamattomaksi jääneitä ei onnistuttu nytkään tavoittamaan. Osalla kansalaisista on siis ”pysyvä” tapa tai taipumus välttää tutkimuksiin osallistuminen.

Tämä kaksiteräinen tulos vastauskadon pysyvyydestä ja muutettavuudesta hankaloittaa survey-tutkijan arkea. Kun toisaalta osa vastauskadosta on ”sittenkin taivutettavissa” mukaan tutkimukseen, herää kysymys, mikä osa ja millä keinoin? Toisaalta osa vastauskadosta on selvästi menetettyjä tapauksia – ainakin nyt käytetyillä menetelmillä – eikä heihin ehkä kannata suunnata panoksia ilman, että tiedonkeruun toteuttamistapaan tehdään muutoksia.

### *Ylipeittomahdollisuus tulisi tarkistaa katoanalyysin yhteydessä*

Laatututkimuksessa ylipeitto oli jopa 4 prosenttia, kun se yleensä on surveyssä noin 1–2 prosenttia. Lähes kaikki tutkimuksen ylipeittotapaukset olivat vuotta aikaisemmin tavoittamattomia. Kertooko tämä siitä, että pieni osa tavoittamattomaksi jääneistä on jo ”liukumassa” ylipeiton suuntaan eli poistumassa tutkimuksen kohdejoukosta? Vaikka ylipeiton kokonaisvaikutus surveyn laatuun on pieni, tulisi ylipeittomahdollisuus tarkistaa katoanalyysin yhteydessä, varsinkin, jos tutkimuksen kentällä oloaika on pitkä.

### *Vastauskatoon kuuluvat eivät huonommassa asemassa*

Vaikka AKU-laaturutkimuksen ja AKU 2006:n vastausosuuksissa on joitakin eroja, niin päätulos oli, etteivät vastaamattomat (vuonna 2006 vastauskatoon jääneet) juurikaan poikenneet demografisilta ominaisuuksiltaan vastanneista. Laaturutkimuksen tulokset eivät tue ajatusta vastauskatoon jääneiden huonommasta yhteiskunnallisesta asemasta ja vähäisemmästä yhteiskunnallisesta aktiivisuudesta. Tulokset olivat jopa päinvastaisia.

AKU 2006:n vastauskatoon jäi ja laaturutkimuksella kiinni saatiin hiukan keskimääräistä enemmän palkansaaajia ja yksityisellä sektorilla työskenteleviä. Erityisesti he osallistuivat lyhyeen, puhelimesta tehtyyn haastatteluun. Onko niin, että kiire ajaa kieltäytymään tutkimuksista ja periksi annetaan vasta, kun tarjolla on lyhyt, haastattelun riisuttu versio?

Laaturutkimukseen osallistuneilla oli myös aavistuksen parempi työmarkkina-asema ja he käyttivät Internetiä aktiivisemmin. Toisaalta osa tässä ilmenneistä pienistä eroista laaturutkimuksen ja AKU 2006:n välillä voi olla heijastusta yhteiskunnallisista muutoksista vuoden aikana tai ne voivat liittyä laaturutkimuksen nuorempaan vastaajajoukkoon.

Esimerkiksi tulokset työttömyyskokemuksista ovat yhteneväisiä työttömyyden kehityksen kanssa. Koko maan työttömyysaste pieneni lähes prosenttiyksikön (7,8 prosentista 6,9 prosenttiin) vuosien 2006 ja 2007 välillä. Määränä tämä tarkoittaa noin 21 000 työtöntä vähemmän vuonna 2007 kuin vuotta aikaisemmin. (<http://tilastokeskus.fi/til/tyti>)

Logit-mallituksen avulla AKU 2006:n ja laaturutkimuksen vastaajajoukko vakioitiin tiettyjen taustamuuttujien (ikä, koulutus, äidinkieli) suhteen. Tällöin, kun laaturutkimus tulkitaan tutkimuskadoksi, oli riski jäädä tutkimuksen ulkopuolella suurempi pysyvän työsuhteen omaavilla, yksityissektoreilla työskentelevillä sekä palkansaaajilla ja yrittäjillä. Tämä vahvistaa tulkintaa, ettei tutkimuskatoon jää vain selkeästi yhteiskunnallisesti huonommassa asemassa olevia (esimerkiksi työttömiä ja määräaikaaisessa työsuhteessa työskenteleviä).

### *Suhtautuminen survey-tutkimuksiin positiivista*

Kun me tutkijat työksemme ”vaivaamme” ihmisiä, oli ilo havaita, että kansalaiset olivat haastattelun päätyttyä tyytyväisiä. Jopa 85 prosenttia ”*sittenkin osallistuneista*” näki tutkimuksilla saatavan arvokasta tietoa ja vain joka neljäs heistä piti osallistumista liian aikaa vievänä. Mitä enemmän vastaaja itse uhrasi aikaansa, sitä harvemmin hän koki tutkimukset rasituksiksi. Mukaan saadut olivat siis mukana survey-tutkimusten ”juonessa”. Osallistumispäätöksen takaa haastattelijoiden ammattimainen ja hyvä suostuttelu sekä saatekirjeet.

### *Tutkimusaineisto laadukasta*

Saatuja laatututkimuksen tuloksia voidaan tietyin varauksin verrata Aikuiskoulutustutkimus 2006 jakaumiin. Tulos aikuiskoulutustutkimuksen keskeisen indikaattorin eli aikuiskoulutukseen osallistumisen osalta oli viehättävä: tutkimuksesta kertaalleen vastaukseen jääneet osallistuvat aikuiskoulutukseen samassa määrin kuin tutkimukseen vuotta aikaisemmin oikeasti osallistuneet. Aikuiskoulutukseen osallistuneiden osuus ei muuttunut, kun laatututkimuksessa vakioitiin ikä ja koulutustausta (vakioidut osallistumisosuudet olivat 51 ja 54 prosenttia). Raportoitu koulutuksiin osallistuminen ei riippunut tiedonkeruutavasta eikä tiedonkeruun ajasta.

Ainoa ero oli, että lyhyessä haastattelussa oli työpaikalla erityisen kouluttajan antamaa perehdytyskoulutusta saaneita enemmän kuin käyntihaastatteluissa. Mahdolliseen ylipäätöksiin löytyy helposti kaksi selitystä. Ensinnäkin perehdytyskoulutus on määritelmänä vaikea<sup>16</sup> ja saattaa olla, ettei koulutuksen sisällön selittämiseen nopeatempoisessa puhelinhaastattelussa (lyhyt haastattelu) ollut mahdollista panostaa riittävästi. Toiseksi tämä koulutus oli puhelinhaastattelussa koulutuskysymyksen kolmantena alakohtana, kun se käyntihaastattelussa tuli vasta kysymysblokin lopussa (ollen 31. alakohta).

Sen sijaan koulutustarvetta nämä alkuperäisen tutkimuksen ulkopuolelle jääneet tai jättäytyneet kokivat hiukan enemmän kuin AKU 2006:ssa todettiin. Lopputulos

---

<sup>16</sup> Perehdytyskoulutuksella tarkoitetaan työpaikalla työn yhteydessä tehtäviin perehdyttävää tai taitoja lisäävää opastusta erityiseltä ohjaajalta, kouluttajalta tai asiantuntijalta.

koulutustarpeesta ei kuitenkaan olisi ollut erilainen, vaikka he olisivat osallistuneet tutkimukseen jo vuonna 2006.

Vastaavanlaisia tuloksia vastauskadon pienestä vaikutuksesta tutkimustuloksiin on saatu muuallakin (Andresen et al. 2008).

Tulos aikuiskoulutukseen osallistumisesta samassa määrin oli sikäli yllättävä, että vastaajan alhainen koulutusaste indikoi suurempaa todennäköisyyttä tutkimuskatoon jäämiseksi. Koulutusaste korreloi vastaavasti myös aikuiskoulutukseen osallistumisen kanssa. Näin olleen olisi ollut oletettua, että vuonna 2006 vastauskatoon jääneet (ja laatututkimukseen osallistuneet) olisivat osallistuneet aikuiskoulutukseen heikommin kuin vuonna 2006 vastanneet. Toisaalta sekä laatututkimukseen että AKU 2006:een osallistui samassa määrin vain perusasteen koulutuksen omaavia.

### *Painokertoimet toimivat*

Saadut tulokset antavat turvaa painokertoimien käytölle. Nehän perustuvat filosofiaan, että osallistumatta jättäneet vastaavat lähes samalla tapaa kuin samanikäiset ja samoilla seuduilla asuvat miehet ja naiset (Tilastokeskuksessa painotus tehdään yleensä iän, sukupuolen ja alueen mukaan). Näin yhden ihmisen vastaus on yleistettävissä tietyin ehdoin edustamaan usean kanssasisarensa tai -veljensä näkemystä.

Laatututkimuksen tulos antaa tosin aihetta pohtia vakavasti myös vastaajan koulutustaustan huomioimista painokertoimia laskettaessa.

## *5.2 Tulosten suhde teoriaan*

Saatuja tuloksia voidaan peilata myös teoriaan kansalaisten osallistumismotiiveista. Tutkimustulokset osoittavat, että jos tutkimukseen liittyy kahdenlaisia lomakkeita, normaali pitkä lomake ja sen jonkin verran lyhyempi versio, kannattaa lyhyttä tarjota vasta, kun henkilö on kieltäytynyt pitkästä. Todennäköisyys, että hän vastaa lyhyempään tavallaan vastavuoroisesti tinkien itsekkin omista päätöksistään on suurempi kuin jos hän saisi suoraan valita, vastaako pitkään vai lyhyeen lomakkeeseen vai ei lainkaan. (Groves et al. 1992.)

Tätä vastavuoroisuuden olettamusta AKU-laaturutkimuksen kokemukset ja tulokset tukivat selkeästi. Laaturutkimuksessa ei vastaajille kerrottu saatekirjeessä lainkaan puhelinhaastattelumahdollisuudesta. Lyhyempää puhelimessa tehtyä haastattelua tarjottiin vastaajille vasta, kun he olivat osoittaneet kiinnostuksen puutetta käyntihaastatteluun. Selvä enemmistö tutkimukseen osallistuneista (66 prosenttia) osallistui puhelinhaastatteluun ja joka kymmenes heltyi vasta viimeisenä vaihtoehtona tarjottuun katohaastatteluun.

Myös johdonmukaisuuden oletama sai osittain tukea tutkimuksesta: laaturutkimuksen vastauskato oli hyvin samanlaista kuin AKU 2006:n vastauskato. Kieltäytyjä kieltäytyivät ja tavoittamattomina olleet jäivät tavoittamatta.

Toisaalta kerran kieltäytyneistä 59 prosenttia antoi jotain tietoa. ”*Ai, tää on näin tärkeää...*” totesi moni osallistuja haastatteliijoille. Tavallaan he tekivät tuolla lausahduksellaan selkeästi näkyväksi oman toimintatapansa muuttamisen (ja epäjohdonmukaiselle käytökselle luotiin julkisesti perustelu).

Käsitykset tutkimusten rasittavuudesta ja jopa hyödyllisyydestä poikkesivat oleellisesti vastaajan vastaustaakan mukaan – yllättäen kylläkin siten, että mitä enemmän vastaaja itse uhrasi laaturutkimukseen aikaa, sitä harvemmin hän koki tutkimukset aikaa vieviksi tai liian henkilökohtaisuuksiin meneviksi. Tämäkin on tulkittavissa johdonmukaisuudeksi: ”eihän puolen tunnin haastattelu oikeastaan kovin rankka suoritus ole”.

Laaturutkimukseen osallistuneilta kysyttiin heidän käsityksiään Tilastokeskuksesta. Useammalle kuin neljälle viidestä laaturutkimukseen osallistuneelle oli Tilastokeskus ennestään tuttu ja 71 prosenttia vastaajista piti virastoa tarpeellisena koko yhteiskunnan kannalta. Vaikka emme tiedä, kuinka muut ajattelivat (laaturutkimuksen vastauskato, aikuiskoulutustutkimuksen se osa vastauskatoa, johon ei otettu uudestaan yhteyttä, AKU 2006 vastanneet) voidaan tuloksien nähdä tukevan ajatusta auktoriteetin vaikutuksesta osallistumishalukkuuteen. Joka kymmenes pitkään haastatteluun suostuneista ilmaisi osallistumispäätökseensä<sup>17</sup> vaikuttaneen sen, että tutkimuksen teki valtion laitos.

---

<sup>17</sup> Jokainen sai nimetä kolme osallistumispäätökseensä vaikuttavaa seikkaa.

Myös tutkimuksen kiinnostavuus ja haastattelijan rooli olivat keskeisiä henkilöiden tehdessä päätöksiä tutkimukseen osallistumisesta.

Laatututkimuksen ja AKU 2006:n saatekirjeet olivat erilaisia. Toisin kuin aikuiskoulutustutkimuksella, oli laatututkimuksella kaksi selkeää aihetta: aikuiskoulutus ja tilastoinnin kehittäminen. Vastaajille kerrottiin kyseessä olevan jatkotutkimus, ja osallistumalla vastaaja on apuna suomalaisen tilastoinnin perustana olevien haastattelututkimuksien kehittämisessä.

Toisena erona ensi kontaktissa oli se, että vuoden 2007 otokseen kuuluville lähetettiin kuvaus keskeisistä aikuiskoulutusmuodoista puhelinhaastattelua helpottamaan. On mahdollista, että tämä A4-dokumentti auttoi vastaajia mieltämään tutkimuksen aihealueen (aikuiskoulutuksen) paremmin, ja tällä voi epäilemättä olla vaikutusta osallistumispäätökseen. Aikuiskoulutus-termin avaaminen ei kuitenkaan näyttäisi vaikuttaneen saatuihin tutkimustuloksiin, mikäli se on tässä esitetyin keinoin arvioitavissa. AKU-laatututkimukseen osallistuneet osallistuivat aikuiskoulutukseen samassa määrin kuin vuotta aikaisemmin varsinaiseen survey-tutkimukseen osallistuneet.

Yllättävää ei ole, että henkilöt, joita ei tavoitettu vuonna 2006 aikuiskoulutustutkimuksen yhteydessä, muistivat vuoden kuluttua ”yhteydenoton” (esimerkiksi saatekirjeen) huonommin kuin vuonna 2006 tutkimuksesta kieltäytyneet. Tosin lähes joka toinen vuonna 2006 tavoittamattomaksi jäänyt, mutta vuotta myöhemmin tutkimukseen suostunut, muisti aikuiskoulutustutkimuksen yhteydenoton (haastattelijan arvion mukaan). Tämä viittaa siihen, että ainakin osaa tavoittamattomiksi luokitelluista henkilöistä voidaan pitää ”piileksijöinä”. Piileksijän tavoitteena mitä ilmeisimmin on välttyä tutkimukseen osallistumiselta ollen koko ajan tavoittamattomissa. Antaako tämä aihetta katokoodien ja kontaktitapojen uudelleen tarkasteluun?

Kummassakaan tutkimuksessa ei käytetty palkkioita. Laatututkimukseen vastanneilta (eli aikuiskoulutustutkimuksen vastauskatoon kuuluvilta) kysyttiin kuitenkin heidän käsitystään palkkion merkityksestä. Mielipiteet vaihtelivat, kuten aikaisemmat tutkimukset ovat antaneet olettaa.

Laatututkimuksessa ei tilastoitu yhteydenottojen määrää, eikä sen vaikutusta tutkimukseen osallistumiseen voida arvioida. Koska laatututkimuksen haastattelut tekivät samat Tilastokeskuksen tilastohaastattelijat, jotka hoitivat myös AKU 2006 haastattelut, ja koska tiedonkeruuprosessiin ei tältä osin tehty muutoksia, voidaan olettaa tavoittamiskertojen olleen suurin piirtein samansuuruisia näissä tutkimuksissa.

Innottomuus osallistua haastatteluun ei viittaa yleiseen yhteiskunnalliseen epäaktiivisuuteen, jos sitä mitataan äänestämiskaudella. Tässä laatututkimuksen tulos on jopa ristiriidassa tutkimukseen osallistumisoletuksien ja aikaisempien tutkimustulosten kanssa. Laatututkimukseen osallistuneet ilmoittivat äänestäneensä vähintäänkin yhtä vilkkaasti kuin kansa keskimäärin.

Survey-asenteita koskevissa kysymyksissä katohaastatteluun osallistuneiden vastaukset poikkesivat muuhun laatututkimuksen osallistuneiden vastauksista. Tämä vahvistaa käsitystä siitä, että tutkimuksen ulkopuolelle jääminen perustellaan (itselle) mielikuvilla tutkimuksen rasittavuudesta ja liiallisesta henkilökohtaisuudesta ja ehkä osittain myös tutkimusten tarpeettomuudella. Mukaan lupautuneet havaitsivat rasituksen loppujen lopuksi kovin pieneksi.

Tätä tukee myös haastatteliijoilta saatu palaute lukuisista Tilastokeskuksessa toteutetuista haastattelututkimuksista. Pitkässäkin haastattelussa haastattelijalla onnistuu pitämään tunnelman hyvänä, ja useimmiten osallistujat ovat haastattelun jälkeen tyytyväisiä. Esimerkiksi AKU 2006 haastatteliijoille suunnatun palautekyselyn mukaan runsaalle kahdelle kolmasosalle haastatteliijoista jäi tutkimuksesta kokonaisuutena myönteinen mielikuva. Näin tuskin olisi, jos haastattelutilanne koettaisiin (haastattelijan ja haastateltavan) pääsääntöisesti rasittavaksi.

Tuloksia voidaan myös tulkita kääntäen kolikko ympäri (eli muutetaan syy-seuraussuhdetta): tutkimukseen mukaan lupautuneet vähättelevät tutkimuksen aiheuttamaa rasitusta halutessaan nähdä itsensä rationaalisesti toimivina henkilöinä.

On muistettava, että koko laatututkimuksen otos on AKU 2006:n vastauskatoa ja siksi positiivinen tulos suhtautumisessa tutkimuksen rasittavuuteen on huomion arvoista.

Kun tuloksia tulkitaan *Leverage-Saliency*-teorian kautta, niin nähdään, että monet erilaiset voimat vaikuttivat henkilöiden tehdessä osallistumispäätöstään. Lähes 60 prosenttia jo kertaalleen vastauskatoon jääneistä suostui yhteistyöhön, ja suurin osa heistä vaihtoehtoisella tiedonkeruutavalla.

### 5.3 Riippuuko saalis haavista?

Kaikki laatututkimuksella saadut jakaumat eivät menneet yksi yhteen Aikuiskoulutustutkimus 2006 jakaumien kanssa, vaan joitakin pieniä poikkeamia ilmeni. Erot voivat johtua tiedonkeruujasta, -menetelmästä tai vastaajajoukosta. Esimerkkinä ajallista eroista voisi olla osan työelämään liittyvien vastausten vaihteluista. Laatututkimuksessa raportoitiin hiukan vähemmän työttömyyskokemuksia, kuten työttömyyslukujen vuosikeskiarvot vuosilta 2006 ja 2007 antavat olettaa. Sen sijaan esimerkiksi aikuiskoulutukseen osallistuminen ei muutu kovin nopeasti ajassa, eikä osuuksissa näkynyt vastaavia eroja laatututkimuksen ja AKU 2006:n kesken.

Eroja voi aiheuttaa myös erilaiset tiedonkeruumenetelmät. Tällaiseksi tulkitsisin sen, että lyhyeen haastatteluun osallistuneet ilmaisivat muita yleisimmin osallistuneensa työhön liittyvään perehdytyskoulutukseen. Kyseinen koulutusmuoto on hankalasti määriteltävissä ja myös sen sijainti kysymyspatteristossa (puhelinhaastattelussa alussa, käyntihaastattelussa lopussa) voivat yhdessä aiheuttaa yliraportointia. Myös ulkoisille ärsykeille herkissä mielipideväittämissä ilmeni joitakin pieniä eroja laatututkimuksen ja AKU 2006 tuloksien välillä (tuloksia ei raportoitu tässä, Pohjanpää 2009a).

Yksi keskeinen tuloksiin vaikuttava seikka on otoksen poimintaperusteet. Laatututkimukseen valittiin mukaan vain osa AKU 2006:n vastauskadosta, ulos rajattiin vuonna 2006 tiukasti kieltäytyneet. Perusteluina oli se, että haluttiin kunnioittaa kieltäytymispäätöstä sekä siten turvata haastattelijoiden työrauha. Asetelmaa, jossa koko vastauskato otettaisiin uudestaan käsittelyyn, pidettiin



tutkimuseettisesti arveluttavana. Kun olen nähnyt laatututkimuksen saaman positiivisen vastaanoton niin haastateltavien kuin haastattelijoiden keskuudessa (mikä näkyy muun muassa korkeana vastausosuutena), pidän mahdollisena kohdistaa katotutkimus tietyin ehdoin koko katopopulaatiolle. Mukaan tutkimukseen kannattaisi ainakin ottaa ne kieltäytyneet, jotka eivät maininneet erityistä syytä kieltäytymiselle.

On yleisesti oletettua, että tutkimuksen vastauskatoon jääneet poikkeavat mukaan saaduista. Tähän ei laatututkimus kuitenkaan antanut kovinkaan paljon vahvistusta siltä osalta vastauskatoa, jota pääsimme kurkistamaan. On muistettava, että meiltä jäi vielä nappaamatta 41 prosenttia vastauskadosta Etelä- ja Länsi-Suomen lääneissä. Kaiken kaikkiaan 420 henkilöä kieltäytyi tai jäi tavoittamattomaksi laatututkimuksessakin ollen siis vastauskatoa kahtena perättäisenä vuotena. Keitä he ovat? Voimme vain toivoa – edelleen – että samantapaisia kuin tutkittavaksi uhrautuneet kansalaiset.

On toki naivia tulkita aineistoja siten, että eroille jakaumissa pyrkii löytämään selityksen, mutta saman oloiset jakaumat hyväksytään ikään kuin ”totuutena”, vailla systemaattista mittausvirheitä olevana. Jakaumathan voivat näyttäytyä samanlaisilta juuri siksi, että mittausvirheet, menetelmäerot ja tutkimusajankohtaan liittyvät erot ”tasoittavat” tuloksen. Nämä analyysit rajasin tutkimuksen ulkopuolelle.

### *Lopuksi*

Laatututkimuksen tulos kertoo, että kun vastauskato ei ole erityisen vino mitattavissa olevien taustamuuttujien suhteen, niin tuloksetkaan eivät ole täysin pielessä – ja tilastoammattilainen voi nukkua yönsä hyvin.

Vaikka vastauskato tässä nimenomaisessa tapauksessa ei vääristänyt tuloksia, niin on edelleen syytä olla varuillaan ja huolissaan erityisesti yksikkökadosta. Vastauskatoon jääneiden puuttuvat vastaukset voivat olla muuta kuin saadut vastaukset jollakin toisella tutkimusasetelmalla ja toisena ajankohtana. Erityisesti tavoittamattomaksi jääneiden yhä kasvava joukko on survey-tutkimuksen laadulle suuri uhka.

Vastauskato heikentää aineiston laatua ja tiedonkeruuprosessin tehokkuutta monella tavoin. Esimerkiksi vastauskadon aiheuttama aineiston pieneneminen on ongelma. Se

suurentaa laskettavien estimaattien keskivirhettä ja heikentää ja joskus jopa estää osa-aineistolle tehtäviä analyyskejä. Lisäksi vastauskato tarkoittaa aina ”hukkaan heitettyjä” resursseja, sillä katotapaustenkin osalta on tehty työtä (vrt. kontaktiyritykset ja motivointiin uhrattua aika). Tälle työlle ei tule varsinaista vastinetta (dataa), ja projektin kokonaistehokkuus (tietoyksiköt per työtunnit) heikkenee.

Myös laatututkimuksessa törmättiin aineiston pienuuteen. Pitkään haastatteluun osallistuneiden pieni määrä johti siihen, että tutkimuksen ulkopuolelle jäi keskeinen ja tärkeä analyysivaihe. Tuloksia ei ollut mahdollista verrata AKU 2006 tuloksiin vastaajien taustatietojen (kuten esimerkiksi ikä, sukupuoli, asuinalue ja koulutus) mukaisissa ryhmissä.

Tämän tutkimuksen ulkopuolelle jäi kaksi ryhmää: (a) ensinnäkin henkilöt, jotka kuuluivat vastauskatoon sekä vuonna 2006 että 2007 ja (b) toiseksi henkilöt, jotka olivat vastauskatoa vuonna 2006, mutta joihin ei otettu uudestaan yhteyttä vuonna 2007. Näistä ensin mainittua ryhmää tarkasteltiin laatututkimuksen katoanalyysin yhteydessä. Jälkimmäinen ryhmä rajattiin jo tutkimusasetelmaa suunniteltaessa ulos. Jatkotutkimuksen osalta olisi tärkeää arvioida, kuinka tutkimustuloksiin vaikuttaa se, että osa otokseen valituista kieltäytyy tutkimuksesta. Kieltäytyneiden vastauksien tutkiminen on sekä teknisesti että tutkimuseettisesti hankalaa.

Vastauskato on mielenkiintoinen, valitettavasti pysyväksi jäänyt ilmiö survey-tutkimusmaailman kentässä. Tarvitaan lisää tutkimusta siitä, miksi henkilöt jäävät vastauskatoon ja poikkeavatko he ja kuinka vastanneista. Kaikki tarvittava tieto ei ole saatavissa rekistereistä.

Olen huolissani siitä, ettei tiedonkeruu näy riittävästi yhtenä keskeisenä tehtäväalueena Tilastokeskuksen julkikuvassa. Olemme tilanteessa, jossa yksittäiset tutkimukset ja tilastot joutuvat lähes akrobaattisiin temppuihin suostutellessaan vastaajia mukaan tutkimukseen. Viranomaiselta tullut kirje ja tutkimuksesta kertova esite eivät vetoa talkoohenkeen, jos ”tilastotalkoista” ei ole ennen tätä yhteydenottoa minkäänlaista tietoa.

Tiedonkeruuprosessiin tulisi kiinnittää entistä enemmän huomiota ja sitä tulisi viilata paremmin eri väestöryhmiä palvelevaksi. Koska vastaajien motiivit tehdä päätöksiä tutkimukseen osallistumisesta vaihtelevat, tulee myös tavoittelutaktiikkojen olla erilaisia. Monimenetelmällisyys, menetelmäeroihin liittyvistä haasteista huolimatta, on lunastanut paikkansa survey-tiedonkeruun kentällä.

## Lähteet

- Abraham, Katharine G., Maitland, Aaron & Bianchi, Suzanne M. (2006): Nonresponse in the American Time Use Survey: Who Is Missing from the Data and How Much Does It Matter? *Public Opinion Quarterly* 70, 676–703.
- Adua, Lazarus & Sharp, Jeff S. (2010): Examining Survey Participation and Response Quality: The Significance of Topic Salience and Incentives. *Survey Methodology* 36, 95–109.
- Allison, Paul D. (2009): *Logistic Regression Using SAS. Theory and Application*. 8th edition. USA: Wiley and SAS Institute.
- Alwin, Duane F. (1991): Research on Survey Quality. *Sociological Methods & Research* 20 (1), 3–29.
- Andresen, Elena M., Machuga, C. Renea, van Booven, Mary Ellen, Egel, John, Chibnall, John T. & Tait, Raymond C. (2008): Effects and Costs of Tracing Strategies on Nonresponse Bias in an Survey of Workers with Low-Back Injury. *Public Opinion Quarterly* 72, 40–54.
- Barón, Juan D., Breunig, Robert V., Cobb-Clark, Deborah, Gørgens, Tue & Sartbayeva, Anastasia (2009): Does the Effect of Incentive Payments on Survey Response Rates Differ by Income Support History? *Journal of Official Statistics* 25 (4), 483–507.
- Biemer, Paul P. & Lyberg, Lars E. (2003): *Introduction to Survey Quality*. Hoboken, New Jersey: A John Wiley & sons Publication.
- Brener, Nancy D., Eaton, Danice K., Kann, Laura, Grunbaum, Jo Anne, Gross, Lori A., Kyle, Tonja M. & Ross, James G. (2006): The Association of Survey Setting and Mode with Self-Reported Health Risk Behaviours Among High School Students. *Public Opinion Quarterly* 70 (3), 354–374.
- Brennan, Mike & Charbonneau, Jan (2009): Improving Mail Survey Response Rates Using Chocolate and Replacement Questionnaires. *Public Opinion Quarterly* 73 (2), 368–378.
- Brennan, Mike & Hoek, Janet (1992): The Behavior of Respondents, Nonrespondents, and Refusers Across Mail Surveys. *Public Opinion Quarterly* 56, 530–535.
- Curtin, Richard, Presser, Stanley & Singer, Eleanor (2000): The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly* 64, 413–428.
- Daniel, Wayne W. (1975): Nonresponse in Sociological Surveys. A Review of Some Methods for Handling the Problem. *Sociological Methods & Research* 3 (3), 291–307.
- Davern, Michael, Rockwood, Todd H., Sherrod, Randy & Cambell, Stephen (2003): Prepaid Monetary Incentives and Data Quality in Face-to-Face Interviews. Data from the 1996 Survey of Income and Program Participation Incentive Experiment. *Public Opinion Quarterly* 67, 139–147.

- De Leeuw, E.D. (1992): *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: T.T. Publikaties.
- (2005): To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics* 21 (2), 233–255.
- (2007): *Innovations in Survey Data Collection: Mixed-mode, Web Surveys, and Visual Design –seminary*. Utrecht University, Holland. 26.–27.4.2007.
- (2008): Choosing the Method of Data Collection. Teoksessa Edith D. de Leeuw, Joop J. Hox and Don A. Dillman (ed.) *International Handbook of Survey Methodology*. New York & London: European Association for Methodology Series. Lawrence Erlbaum Associates, 113–135.
- De Leeuw, Edith D., Callegaro, Mario, Hox, Joop, Korendijk, Elly & Lensvelt-Mulders, Gerty (2007): The Influence of Advance Letters on Response in Telephone Surveys. A Meta-Analysis. *Public Opinion Quarterly* 71 (3), 413–443.
- De Leeuw, Edith D., Hox, Joop J. & Dillman, Don A. (2008): Mixed-mode Surveys: When and Why. Teoksessa Edith D. de Leeuw, Joop J. Hox & Don A. Dillman (ed.): *International Handbook of Survey Methodology*. New York & London: European Association for Methodology Series. Lawrence Erlbaum Associates, 299–316.
- Dillman, Don A. (2007): *Mail and Internet Surveys. The Tailored Design Method, 2007. Update With New Internet, Visual, and Mixed-Mode Guide*. New Jersey: Jon Wiley & Sons Inc.
- Dillman, Don A., Eltinge, John L., Groves, Robert M. & Little, Raderick J. A. (2002): *Survey Nonresponse in Design, Data Collection, and Analysis*. Teoksessa Robert M. Groves, Don A. Dillman, John L. Eltinge & Roderick J. A. Little (eds.) (2002): *Survey Nonresponse*. New York: Wiley.
- Dillman, Don A., Smyth, Jolene D. & Christian, Leah M. (2009): *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method*. Third ed. New Jersey: John Wiley & Sons Inc.
- Fricker, Scott, Galesic, Mirita, Tourangeau, Roger & Yan, Ting (2005): An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly* 69 (3), 370–392.
- Fricker, Scott & Tourangeau, Roger (2010): Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys. *Public Opinion Quarterly* 74 (5), 934–955.
- Gmel, Gerhard (2000): The Effect of Mode of Data Collection and of Non-Response on Reported Alcohol Consumption: a Split-Sample Study in Switzerland. *Addiction* 95 (1), 123–134.
- Granberg, Donald & Homberg, Soren (1991). Self-Reported Turnout and Voter Validation. *American Journal of Political Science*, 35 (2), 448–459.

- Greenfield, Thomas K., Midanik, Lorraine T. & Rogers, John D. (2000): Effects of Telephone Versus Face-to-Face Interview Modes on Reports of Alcohol Consumption. *Addiction* 95, 277–284.
- Groves, Robert M. (1998): *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- (2006): Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70, 646–675.
- Groves, Robert M., Cialdini, Robert B. & Couper, Mick P. (1992): Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly* 56, 475–495.
- Groves, Robert M. & Couper, Mick P. (1998): *Nonresponse in Household Interview Surveys*. New York: Wiley.
- Groves, Robert M., Singer, Eleanor & Corning, Amy (2000): Leverage-Saliency Theory of Survey Participation. Description and an Illustration. *Public Opinion Quarterly* 64, 299–308.
- Groves, Robert M., Presser, Stanley & Dipko, Sarah (2004): The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly* 68 (1), 2–31.
- Groves, Robert M., Couper, Mick P., Presser, Stanley, Singer, Eleanor, Tourangeau, Roger, Acosta, Giorgia Piani & Nelson, Lindsay (2006): Experiments in Producing Nonresponse Bias. *Public Opinion Quarterly* 70 (5), 720–736.
- Heerwegh, Dirk (2005): Effects of Personal Salutations in e-mail Invitations to Participate in a Web Survey. *Public Opinion Quarterly* 69 (4), 588–598.
- Hembroff, Larry, Rusz, Debra, Rafferty, Ann, McGee, Harry & Ehrlich, Nathaniel (2005): The Cost-Effectiveness of Alternative Advance Mailings in a Telephone Survey. *Public Opinion Quarterly* 69 (2), 232–245.
- Hosmer, David & Lemeshow, Stanley (2000): *Applied Logistic Regression*. 2nd edition. New York: Wiley.
- Isotalo, Jarkko (2012): Johdatus yleistettyihin lineaarisiin malleihin. Informaatiotieteiden yksikön raportteja 8/2012. Tampere: Tampereen yliopisto.
- Jäckle, Annette & Lynn, Peter (2008): Respondent Incentives in a Multi-Mode Panel Survey: Cumulative Effects on Nonresponse and Bias. *Survey Methodology* 34 (1), 105–117.
- Keeter, Scott, Miller, Catherine, Kohut, Andrew, Groves, Robert M. & Presser, Stanley (2000): Consequences of Reducing Nonresponse in a National Telephone Survey. *Public Opinion Quarterly* 64:125–48.
- Keeter, Scott, Kennedy, Courtney, Dimock, Michael, Best, Jonathan & Craighill, Peyton (2006): Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey. *Public Opinion Quarterly* 70, 759–79.
- Keinänen, Päivi (2009, toim.): *Viisi vuosikymmentä työvoiman tilastointia. Työvoimatutkimuksen historiakatsaus*. Helsinki: Tilastokeskus.

- Kuusela, Vesa (2009): Otostutkimukset otettiin hitaasti käyttöön virallisissa tilastoissa. *Hyvinvointikatsaus* (2), 50–53.
- Laaksonen, Seppo (2010): *Surveymetodiikka*. [www.bookboon.com](http://www.bookboon.com)
- Lehtonen, Risto & Pahkinen, Erkki (2004): *Practical Methods for Design and Analysis of Complex Surveys* (2nd. Ed.) Chichester: John Wiley and Sons.
- Lin, I.-Fen & Scaffer, Nora C. (1995): Using Survey Participants to Estimate the Impact of Nonparticipation. *Public Opinion Quarterly* 59, 236–258.
- Link, Michael W. & Mokdad, Ali (2005): Advance Letter as a Means of Improving Respondent Cooperation in random Digit Dial Studies. *Public Opinion Quarterly* 69 (4), 572–587.
- Losch, Mary E., Maitland, Aaron, Lutz, Gene, Mariolis, Peter & Gleason, Steven C. (2002): The Effect of Time of Year of data Collection on Sample Efficiency. An Analysis of Behavioral Risk Factor Surveillance Survey. *Public Opinion Quarterly* 66, 594–607.
- Lynn, Peter (2008): The Problem of Nonresponse. Teoksessa Edith D. de Leeuw, Joop J. Hox & Don A. Dillman (ed.): *International Handbook of Survey Methodology*. New York & London: European Association for Methodology Series. Lawrence Erlbaum Associates, 35–55.
- Lynn, Peter, Clarke, Paul, Martin, Jean & Sturgis, Patrick (2002): The Effects of Extended Interviewer Efforts on Nonresponse Bias. Teoksessa Groves, Robert M., Dillman, Don A., Eltinge, John L. & Little, Roderick J. A. (eds.): *Survey Nonresponse*. New York: John Wiley and Sons, 135–148.
- Macer, Tim (2003): The Next Challenge: Efficient and Effective Mixed- and Multi-Mode Research. [http://www.meaning.uk.com/resources/articles\\_papers/files/dutch\\_mra\\_2003.pdf](http://www.meaning.uk.com/resources/articles_papers/files/dutch_mra_2003.pdf)
- Mann, Christopher B. (2005): Do Advance Letters Improve Preelection Forecast Accuracy? *Public Opinion Quarterly* 69 (4), 561–571.
- Mariano, Louis T. & Kadane, Joseph B. (2001): The Effect of Intensity of Effort to Reach Survey Respondents: A Toronto Smoking Survey. *Methodology* 2001 (2), 131–142.
- Menard, Scott (2002): *Applied Logistic Regression Analysis*. 2nd edition. Series: Quantitative Applications in the Social Sciences n:o 106. Thousand Oaks, London & New Delhi: Sage Publications.
- Minska bortfallet (1997). Japac, Lilli, Ahtiainen, Antti, Hörngren, Jan, Lindén, Håkan, Lyberg, Lars & Nilsson, Per (toim.). Sverige: Statistiska centralbyrån. [http://www.scb.se/statistik/\\_publikationer/ov9999\\_2000i02\\_br\\_x97%C3%96p9701.pdf](http://www.scb.se/statistik/_publikationer/ov9999_2000i02_br_x97%C3%96p9701.pdf)
- Mowen, John C. & Cialdini, Robert B. (1980): On Implementing the Door-in-the-Face-Compliance Technique in a Business Context. *Journal of Marketing Research* 17, 253–258.
- Pahkinen, Erkki & Lehtonen, Risto (1989): *Otanta-asetelmat ja tilastollinen analyysi*. Helsinki: Gaudeamus.

- Peytcheva, Emilia & Groves, Robert M. (2009): Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates. *Journal of Official Statistics* 25 (2), 193–201.
- Peytchiev, Andy, Baxter, Rodney K. & Carley-Baxter, Lisa R. (2009): Not All Survey Effort Is Equal. Reduction of Nonresponse Bias and Nonresponse Error. *Public Opinion Quarterly* 73 (4), 785–806.
- Pohjanpää, Kirsti (2009a): Keitä on ne, joita ei haaviin saada? AKU-laatu 2007:n tuloksia. Tilastokeskuksen sisäinen raportti 25.9.2009.
- (2009b): Ajat muuttuvat, muuttuvatko menetelmät? Hyvinvointikatsaus 2009 (4), 48–50.
- (2010): Tilastoinnin merkityksen ymmärtäminen lisää tutkimuksiin osallistumista. Hyvinvointikatsaus 2010 (4), 55–60.
- (2011): Do non-participants answer similarly as survey-participants? – An example of AES-quality-survey. The 4th Conference of the European Survey Research Association (ESRA). Lausanne, Switzerland. 18–22 July 2011. <http://www.europeansurveyresearch.org/conferences/lausanne-2011>
- Pohjanpää, Kirsti, Niemi, Helena & Ruuskanen, Timo (2008): Osallistuminen aikuiskoulutukseen. Aikuiskoulutustutkimus 2006. Koulutus 2008. Helsinki: Tilastokeskus.
- Rita, Hannu (2004): Vetosuhde (odds ratio) ei ole todennäköisyyksien suhde. *Metsätieteen aikakauskirja* 2004 (2), 207–211.
- Romanov, Dmitri & Nir, Michal (2010): Get It or Drop It? Cost-Benefit Analysis of Attempts to Interview in Household Surveys. *Journal of Official Statistics* 26 (1), 165–191.
- Singer, Eleanor, van Hoewyk, John & Maher, Mary P. (2000): Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly* 64, 171–188.
- Teitler, Julien, Reichman, Nancy & Sprachman, Susan (2003): Costs and Benefits of Improving Response Rates for a Hard-to-Reach Population. *Public Opinion Quarterly* 67, 126–138.
- Tourangeau, Roger, Groves, Robert M. & Redline, Cleo D (2010): Sensitive Topics and Reluctant Respondents. Demonstrating a Link Between Nonresponse Bias and Measurement Error. *Public Opinion Quarterly* 74 (3 Fall), 413–432.
- Trussell, Norm & Lavrakas, Paul J. (2004): The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response. Is There an Optimal Amount? *Public Opinion Quarterly* 68 (3 Fall), 349–367.
- Vastauskadon vuosiraportti 2009. Tilastokeskuksen sisäinen raportti.
- Virtanen, Hannu (1989): Survey on Survey. Respondents' experiences about participating in the Household Income Distribution Survey. The 2nd International Workshop on Nonresponse and Survey Participation. Washington, DC, USA. 16.–18.10.1991.



- Voogt, Robert J.J. & Saris, Willem E. (2005): Mixed Mode Designs: Finding the Balance Between Nonresponse Bias and Mode Effects. *Journal of Official Statistics* 21 (3), 367–387.
- Wang, Y., Lee, C., Lew-Ting, C., Hsiao C. K., Chen, D. & Chen, W. J. (2005): Survey of Substance Use Among High School Students in Taipei: Web-Based Questionnaire Versus Paper-and-Pencil Questionnaire. *Journal of Adolescent Health* 37, 289–295.
- Yan, Ting, Curtin, Richard & Jans, Matthew (2010): Trends in Income Nonresponse Over Two Decades. *Journal of Official Statistics* 26 (1), 145–164.

### *Liite 1. Katokysymykset AKU-laatatutkimuksessa*

Koodit kysymyksen perässä viittaavat haastattelumuotoon: PI = pitkä haastattelu (käynti), LY = lyhyt haastattelu (puhelin) ja KA = katohaastattelu (puhelin)

#### **Tilastokeskusta koskevat kysymykset**

**K1.** Oliko Tilastokeskus Teille ennen yhteydenottoamme tuttu? **PI, LY, KA**

Tarkoittaa yhteydenottoa jo vuonna 2006.

- |   |       |
|---|-------|
| 1 | Kyllä |
| 2 | Ei    |

**K2.** Onko Tilastokeskus mielestänne: **PI, LY, KA**

- 1 Tarpeellinen koko yhteiskunnan kannalta
- 2 Hyödyksi vain pienelle osalle suomalaisia
- 3 Tarpeeton laitos?

**K3.** Onko tilastotieto mielestänne tavalliselle kansalaiselle: **PI**

- 1 Helposti ymmärrettävää
- 2 Melko helposti ymmärrettävää
- 3 Melko vaikeasti ymmärrettävää
- 4 Sitä on lähes tai täysin mahdotonta ymmärtää?

#### **Osallistumispäätöstä koskevat kysymykset**

**K4.** Missä vaiheessa päätitte tulla mukaan tähän tutkimukseen? **PI**

Tarkoitetaan juuri tätä vuonna 2007 tehtävää AKU-laadunvarmennustiedonkeruuta.

- 1 Saatuaan ennakkokirjeen
- 2 Haastattelijan ottaessa yhteyttä
- 3 Sen jälkeen, kun sai haastattelijalta lisätietoja tutkimuksesta

**K5.** Mitkä asiat vaikuttivat eniten päätökseenne osallistua tähän Aikuiskoulutus-tutkimukseen? **PI**

Mainitkaa kolme tärkeintä.

#### **VASTAUSKORTTI**

- 1 Sattui olemaan vapaata aikaa
- 2 Minusta on hauska keskustella ihmisten kanssa
- 3 En tohtinut kieltäytyä
- 4 Tutkimuksen teki valtion laitos
- 5 Haastattelija vaikutti ammattitaitoiselta
- 6 Tutkimus vaikutti kiinnostavalta
- 7 Tutkimuksella saadaan tärkeitä tietoja
- 8 Siten voi vaikuttaa asioihin
- 9 Perheenjäsenet, ystävät tai tuttavat kehottivat
- 10 Vastaan yleensä kaikkiin tutkimuksiin
- 11 Tutkimuksiin vastaaminen on jokaisen velvollisuus
- 12 Jokin muu asia, mikä? => K5x

**K5x.** Mikä muu asia vaikutti osallistumiseenne? \_\_\_\_\_ **PI**

**K6.** Ohessa on joukko väittämiä, jotka kuvaavat erilaisia suhtautumistapoja tutkimuksiin. Mitä mieltä Te olette väitteistä. **PI, LY, (KA)**

- 1 Täysin samaa mieltä
- 2 Jokseenkin samaa mieltä
- 3 Ei samaa, mutta ei eri mieltäkään
- 4 Jokseenkin eri mieltä
- 5 Täysin eri mieltä

a Yksityisen kansalaisen antamalla tiedoilla ei ole mitään merkitystä. **PI, LY**

b Tutkimuksissa kysytään liian henkilökohtaisia asioita. **PI, LY, KA**

c Tutkimuksilla saadaan arvokasta tietoa suunnittelua varten. **PI, LY, KA**

d Vastaamalla tutkimukseen yksityinen kansalainen voi vaikuttaa asioihin. **PI, LY**

e Tutkimukseen osallistuminen antaa ajattelemisen aihetta omankin elämän kannalta. **PI, LY**

f Tutkimukseen osallistuminen vie liian paljon aikaa. **PI, LY, KA**

### Palkkiota ja palkitsemista koskevat kysymykset

**K7.** Erityisesti ulkomailla on tutkimuksiin osallistujille joskus annettu pieniä tavara- tai rahapalkintoja. Mitä mieltä olette tällaisesta käytännöstä Suomessa? **PI, LY, KA**

Palkitseminen on

- 1 Erittäin tarpeellista
- 2 Tarpeellista
- 3 Vaikea sanoa
- 4 Melko tarpeetonta
- 5 Täysin tarpeetonta

**K8.** Jos vastaajia jollakin tavalla palkittaisiin, niin mikä olisi Teidän mielestänne sopivin palkkio esimerkiksi tunnin kestävästä haastattelusta? **PI, LY**

- 1 vastanneille maksetaan rahapalkkio => K81
- 2 vastanneille annetaan pieni tavarapalkkio
- 3 Vastanneiden kesken arvotaan muutamia suurempia palkintoja (esim. etelänmatkoja)
- 4 Vastaajat saavat tutkimustuloksista kertovan julkaisun
- 5 Muu palkkio, mikä? => K8x

**K81.** Minkä suuruinen rahasumma olisi mielestänne riittävä? \_\_\_\_\_euroa **PI, LY**

**K8x.** Mikä muu palkkio? \_\_\_\_\_ **PI**

**K9.** Tilastokeskus on valtion laitos, jonka tehtävänä on kerätä yhteiskuntaoloja kuvaavia tietoja. Sopisiko vastaajien palkitseminen mielestänne Tilastokeskuksen haastattelututkimuksiin? **PI, LY**

- 1 Sopisi hyvin
- 2 Saattaisi sopia
- 3 Tuskin sopisi
- 4 Ei sopisi
- 5 EOS

**K10.** Äänestittekö eduskuntavaaleissa maaliskuussa 2007?

**PI, LY, KA**

- 1 Kyllä
- 2 En
- 0 Ei ollut äänioikeutta

*Liite 2. Luettelo aikuiskoulutusmuodoista (lähetetty saatekirjeen mukana)*

# Aikuiskoulutus

Tutkimuksessa kartoitetaan aikuiskoulutukseen osallistumista seuraavanlaisella kysymyksellä.

**AIVAN YHTÄ TÄRKEÄÄ KUIN TIETO KOULUTUKSEEN OSALLISTUMISESTA ON TIETO SIITÄ, ETTEI OLE OSALLISTUNUT MIHINKÄÄN KOULUTUKSEEN VUODEN AIKANA.**

**Oletteko osallistunut Suomessa viimeksi kuluneiden 12 kuukauden aikana seuraaviin opintoihin tai kursseille?**

- Opiskeluksi katsotaan myös keskeytyneet opinnot.
- Jokainen käymänne koulutus tulee merkitä vain yhteen alakohtaan.
- Vain opiskelu, joka on kestänyt yhteensä vähintään 6 tuntia.

1. työnantajan järjestämä koulutus
2. konferenssi tai seminaari
3. työpaikalla työn yhteydessä tehtäviin perehdyttävä tai taitoja lisäävä opastus erityiseltä ohjaajalta, kouluttajalta tai asiantuntijalta
4. kansalais- tai työväenopiston kurssi
5. jonkin järjestön, yhdistyksen tai liiton järjestämä koulutus
6. aikuisille järjestetty koulutus ammatillisessa oppilaitoksessa, ammattikorkeakoulussa tai aikuiskoulutuskeskuksessa
7. pääasiassa nuorille tarkoitettu tutkintoon johtava koulutus ammatillisessa oppilaitoksessa tai ammattikorkeakoulussa
8. opiskelu yliopistossa tai korkeakoulussa
9. erityisen koulutusta antavan koulutusyrityksen tai koulutuskeskuksen kurssi
10. työllisyyskurssi (työvoimaviranomaisten kustantama, työvoimapoliittinen koulutus)
11. oppisopimuskoulutus
12. opiskelu urheiluopistossa, musiikkiopistossa, konservatoriossa tai tanssikoulussa tai -opistossa
13. opiskelu lukiossa, iltalukiossa tai iltakoulussa
14. kansanopiston tai kansankorkeakoulun kurssi
15. kesäyliopiston kurssi
16. jokin muu koulutus Suomessa

*Liite 3. AKU 2006:n katokoodit ja AKU-laatuselvityksen otos*

KATOKOODIT		Laatuselvitys, otos
<b>Kieltäytyy</b>		
11	kieltäytyy, ei erityistä syytä	
12	kieltäytyy ajanpuutteen takia	kyllä
13	kieltäytyy periaatteellisista syistä	
14	kieltäytyy sairauden takia	
16	kieltäytyy tutkimuksen aiheen takia	
17	epäsuora kieltäytyminen	kyllä
18	joku muu kieltää osallistumisen	
19	kieltäytyy tiedonkeruutavan takia	
<b>Ei tavoiteta</b>		
21	ei tavoiteta, osoite ja puh tiedossa	kyllä
22	ei tavoiteta, osoite tuntematon	kyllä
23	ei tavoiteta, poissa tilapäisesti	kyllä
25	ei tavoiteta, laitoshoidossa tilapäisesti	kyllä
<b>Muu syy</b>		
31	muu syy	
32	kielivaikeudet	
<b>Ylipeitto</b>		
51	kuollut	
52	ulkomailla pysyvästi	
53	laitoshoidossa pysyvästi	
58	laitoshoidossa tms., rekisteritieto	
59	kuollut, rekisteritieto	

*Liite 4. AKU-laaturutkimuksen saatekirjeet: ei-tavoitetut ja kieltäytyneet***Arvoisa vastaanottaja**

Tilastokeskuksen tehtävänä on laatia Suomen valtion viralliset tilastot. Vuonna 2006 tehtiin yhdessä opetusministeriön ja Euroopan tilastoviraston (Eurostat) kanssa Aikuiskoulutustutkimus, joka on maamme aikuiskoulutustilastoinnin perusta. Te kuuluitte tutkimuksen satunnaisesti valittuun otokseen, mutta emme tavoittaneet Teitä tuolloin haastattelua varten. Yritämme nyt uudestaan, koska haluamme varmistaa tutkimusaineiston laadukkuuden. On tärkeää, että aineisto kuvaa kattavasti kaikkia suomalaisia.

Toivoisimmekin, että voisimme vielä haastatella Teitä. Antamanne tiedot käsitellään ehdottoman luottamuksellisesti.

**- Kaikkien osallistuminen on erittäin tärkeää kokonaiskuvan saamiseksi -**

Aikuiskoulutustutkimuksessa selvitetään, millaista koulutusta 18 - 66 -vuotiaat suomalaiset ovat saaneet viimeisen vuoden aikana. **Myös tieto siitä, ettei ole osallistunut koulutukseen, on erittäin tärkeää.** Koulutuskokemuksia koskevia tietoja ei ole mahdollista saada muuten kuin kysymällä ihmisiltä itseltään. Alun perin haastateltavat on poimittu satunnaisesti Tilastokeskuksen Suomen väestöä koskevasta tietokannasta. Vuonna 2006 haastattelimme yli 4 300 henkilöä.

Tällä jatkotutkimuksella saadaan tietoa siitä, kuvaako tutkimukseen vuonna 2006 osallistuneet tilastoinnin kannalta riittävän hyvin kaikkia suomalaisia. Osallistumalla autatte meitä kehittämään suomalaisen tilastoinnin perustana olevia haastattelututkimuksia. Tavoitteenamme on tehdä yhä laadukkaampia tilastoja siten, että vastaajia vaivataan mahdollisimman vähän.

**- Tilastotietoa tarvitaan päätöksenteon pohjaksi - suomalaisten hyväksi -**

Tilastohaastattelija (yhteystiedot kirjeen yläkulmassa) ottaa Teihin yhteyttä ja kysyy mahdollisuuttanne osallistua tähän jatkotutkimukseen. Projektin tutkija Kirsti Pohjanpää p. (09) 1734 2604 tai [kirsti.pohjanpaa@tilastokeskus.fi](mailto:kirsti.pohjanpaa@tilastokeskus.fi) antaa mielellään lisätietoja tutkimuksesta. Tietoa löytyy myös [www.tilastokeskus.fi/aikuiskoulutustutkimus](http://www.tilastokeskus.fi/aikuiskoulutustutkimus).

Arvokkaasta avustanne jo etukäteen kiittäen.

Riitta Harala  
Tilastojohtaja



### *Arvoisa vastaanottaja*

Tilastokeskuksen tehtävänä on laatia Suomen valtion viralliset tilastot. Vuonna 2006 tehtiin yhdessä opetusministeriön ja Euroopan tilastoviraston (Eurostat) kanssa Aikuiskoulutustutkimus, joka on maamme aikuiskoulutustilastoinnin perusta. Te kuulitte tutkimuksen satunnaisesti valittuun otokseen, mutta Teillä ei tuolloin ollut mahdollisuutta osallistua. Lähestymme Teitä nyt uudestaan, koska haluamme varmistaa tutkimusaineiston laadukkuuden. Toivoisimmekin, että voisimme vielä haastatella teitä. Antamanne tiedot käsitellään ehdottoman luottamuksellisesti.

#### *- Kaikkien osallistuminen on erittäin tärkeää kokonaiskuvan saamiseksi -*

Aikuiskoulutustutkimuksessa selvitetään, millaista koulutusta 18 - 66 -vuotiaat suomalaiset ovat saaneet viimeisen vuoden aikana. **Myös tieto siitä, ettei ole osallistunut koulutukseen, on erittäin tärkeää.** Koulutuskokemuksia koskevia tietoja ei ole mahdollista saada muuten kuin kysymällä ihmisiltä itseltään. Alun perin haastateltavat on poimittu satunnaisesti Tilastokeskuksen Suomen väestöä koskevasta tietokannasta. Vuonna 2006 haastattelimme yli 4 300 henkilöä.

Tällä jatkotutkimuksella saadaan tietoa siitä, kuvaako tutkimukseen vuonna 2006 osallistuneet tilastoinnin kannalta riittävän hyvin kaikkia suomalaisia. Osallistumalla autatte meitä kehittämään suomalaisen tilastoinnin perustana olevia haastattelututkimuksia. Tavoitteenamme on tehdä yhä laadukkaampia tilastoja siten, että vastaajia vaivataan mahdollisimman vähän.

#### *- Tilastotietoa tarvitaan päätöksenteon pohjaksi - suomalaisten hyväksi -*

Tilastohaastattelija (yhteystiedot kirjeen yläkulmassa) ottaa Teihin yhteyttä ja kysyy mahdollisuuttanne osallistua tähän jatkotutkimukseen. Projektin tutkija Kirsti Pohjanpää p. (09) 1734 2604 tai [kirsti.pohjanpaa@tilastokeskus.fi](mailto:kirsti.pohjanpaa@tilastokeskus.fi) antaa mielellään lisätietoja tutkimuksesta. Tietoa löytyy myös [www.tilastokeskus.fi/aikuiskoulutustutkimus](http://www.tilastokeskus.fi/aikuiskoulutustutkimus).

Arvokkaasta avustanne jo etukäteen kiittäen.

Riitta Harala  
Tilastojohtaja



## Liite 5. Logit-mallitus: muuttujat ja mallit

### A Käytetyt muuttujat

Oheiseen taulukkoon on koottu luvuissa 4.6 ja 4.14 esitettyjen logit-mallituksissa käytettyjä dummy-muuttujat ja niiden luokitukset. Koska osa kysymyksistä oli esimerkiksi työelämään liittyä ja siksi niitä ei ole esitetty kaikille vastaajille, on saattaa havaintojen koko vaihdella muutamien muuttujien osalta. Viimeiseen sarakkeeseen on merkitty poikkea havaintomäärä. Normaalista analysoitavassa aineistossa oli mukana  $n = 3\,793$  havaintoa.

**Liitetaulukko 1.** Logit-mallituksissa käytetyt dummy-muuttujat.

<i>Muuttujan nimi</i>	<i>Muuttujan kuvaus, sisältö</i>	<i>Luokka 0</i>	<i>Luokka 1</i>	<i>Poikkeus n</i>
<b>vakio</b>	vakiotermi			
<b>laatuAKU</b>	osallistuminen laatututkimukseen	kato	vastanneet	
<b>AKU06</b>	osallistuminen AKU 2006	kato	vastanneet	
<b>sukup</b>	sukupuoli	mies	nainen	
<b>koulu1</b>	perusasteen koulutus	ei	kyllä	
<b>koulu3</b>	keskiasteen koulutus	ei	kyllä	
<b>koulu5</b>	korkea-asteen koulutus	ei	kyllä	
<b>ika1</b>	ikäluokka 18–34	ei	kyllä	
<b>ika2</b>	ikäluokka 35–44	ei	kyllä	
<b>ika3</b>	ikäluokka 45–64	ei	kyllä	
<b>kieli</b>	äidinkieli	muut	suomi	
<b>laani</b>	lääni	Etelä-Suomi	Länsi-Suomi	
<b>loptulos06</b>	AKU 2006 kadon syy	kieltäytyi	ei tavoitettu	
<b>A1</b>	palkansaaja + yrittäjä	ei	kyllä	
<b>A5</b>	pysyvä työsuhde	ei	kyllä	2 452
<b>A26</b>	yksityissektori	ei	kyllä	2 651
<b>G2</b>	internetin käyttö	harvemmin kuin päivittäin	päivittäin	
<b>G89</b>	vieraiden kielten käyttö	ei päivittäin	päivittäin	
<b>H2</b>	ollut työttömänä viim. 5 v. aikana	ei	kyllä	3 376
<b>J1</b>	lisäkoulutustarve, työ	ei	kyllä	3 461
<b>J3</b>	lisäkoulutustarve, harrastukset	ei	kyllä	
<b>akuos</b>	aikuiskoulutukseen osallistuminen viim. 12 kk aikana	ei	kyllä	

### **B Esimerkki logit-mallituksesta ja tuloksien tulkinnasta**

Logit-mallituksessa SAS:n proc logistic -ohjema käyttää vertailuluokkana luokkaa 0. Ohjelman laskemat estimaatit ja odds ratio -luvut ovat tulkittavassa seuraavan esimerkin avulla.

Muuttuja	Muuttujan kuvaus, sisältö	Luokka 0	Luokka 1	Estimaatti	OR
sukup	sukupuoli	mies	nainen	<b>0,6931</b>	2,00
laani	lääni	Etelä-Suomi	Länsi-Suomi	-0,4005	0,67

**OR:n laskenta:** Jos tilasto-ohjelma ei tuota OR-arvoja, saadaan ne laskettua estimaateista seuraavasti:

$$OR(sukup) = e^{0,6931} = 2,00 \quad \text{ja} \quad OR(laani) = e^{-0,4005} = 0,67$$

**Saadun tuloksen tulkinta:** Naisten vetosuhde selitettävän ilmiön suhteen on 2,0-kertainen verrattuna miehiin. Koska laani-muuttujan osalta  $OR < 1$ , on tulos helpompi muotoilla sanalliseen muotoon OR:n vastaluvun (vrt. kaava 3.23, sivu 34) ja luokituksen muutoksen avulla.

$OR_{apu} = 1/0,67 = 1,49$ , jolloin voidaan todeta, että eteläsuomalaisten ”riski” tarkasteltavan ilmiön osalta on 1,5-kertainen verrattuna länsisuomalaisten ”riskiin”.

Huomaa, ettei OR ilmaise todennäköisyyksiä (Rita 2004). Vaikka termi vetosuhde tai veto ei aukea suoraan arkikäytössä, niin muistaessaan kaavan (3.23) on helppo luopua puheesta OR:sta todennäköisyyksien kertoimena:

$$OR = \frac{p_i/(1-p_i)}{p_1/(1-p_1)} = \frac{p_i(1-p_1)}{p_1(1-p_i)}$$

## C Analyysissa 1–3 tehdyt mallitukset

### Analyysi 1. Laatututkimukseen osallistuminen

Täysi malli, josta lähdettiin liikkeelle:

```
laatuAKU = sukup + ika2 + ika3 + koulul1 + koulu3 + loptulos06 + laani
+ kieli + sp_ika2 + sp_ika3
```

**Liitetaulukko 2.** Laatututkimukseen osallistuminen (osallistuneet ja vastauskato), logit-mallitus (A ~ B), täysi malli.

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi Sq
Intercept	1	-0.7603	0.3926	3.7503	0.0528
koulul1	1	0.3636	0.1798	4.0904	0.0431
koulu3	1	0.00586	0.1557	0.0014	0.9700
loptulos06	1	0.5853	0.1323	19.5656	<.0001
laani	1	-0.4068	0.1368	8.8473	0.0029
Sukup	1	0.0379	0.1940	0.0382	0.8451
ika2	1	0.7521	0.5253	2.0501	0.1522
ika3	1	0.5676	0.4577	1.5379	0.2149
kieli	1	0.0902	0.2060	0.1917	0.6615
sp_ika2	1	-0.4063	0.3407	1.4217	0.2331
sp_ika3	1	-0.4097	0.3004	1.8601	0.1726

Hyväksytty malli:

```
laatuAKU = vakio + koulul1 + loptulos06 + laani
```

Mallituksen tulokset on esitetty taulukossa 4.8.

Mallin Hosmer & Lemeshow -testin arvot kertovat, että malli on merkitsevä:

Chi-Square	DF	Pr > Chi Sq
1.2376	5	0.9412

Tässä testissä nollahypoteesi  $H_0$  : malli on yhteensopiva aineiston kanssa.

## Analyysi 2. AKU 2006:een osallistuminen ja tutkimuksen vastauskato

AKU 2006:een osallistumista ja vastauskatoa tarkasteltiin hyödyntäen AKU-laaturitkimyksellä saatuja tietoja AKU 2006:n vastauskadosta.

### a) Kaikki muuttujat (havaintoja n = 2 451)

Täysi malli, josta lähdettiin liikkeelle:

```
aku06 = koulu3 + koulu5 + laani + sukup + ika2 + ika3 + sp_ika2 +
sp_ika3 + kieli + g89 + g2 + j3 + A5 + A26 + h2 + j1
```

**Liitetaulukko 3.** AKU 2006:een osallistuminen (osallistuminen ja vastauskato), logit-mallitus ( $C \sim A$  pitkä, lyhyt), täysi malli.

Parameter	DF	Estimate	Standard	Wald	Pr > Chi Sq
			Error	Chi-Square	
Intercept	1	-1.6087	0.4649	11.9744	0.0005
koulu3	1	-0.1722	0.1651	1.0880	0.2969
koulu5	1	-0.5322	0.1815	8.5986	0.0034
Laani	1	0.0537	0.1157	0.2158	0.6423
sukup	1	0.00568	0.1771	0.0010	0.9744
ika2	1	-0.4767	0.4452	1.1467	0.2842
ika3	1	-0.4455	0.4185	1.1330	0.2871
sp_ika2	1	0.2302	0.2854	0.6505	0.4199
sp_ika3	1	-0.0448	0.2710	0.0273	0.8687
kieli	1	-0.5586	0.2051	7.4156	0.0065
G89	1	-0.1810	0.1381	1.7177	0.1900
G2	1	0.1398	0.1450	0.9301	0.3348
J3	1	-0.1958	0.1227	2.5483	0.1104
A5	1	0.4682	0.1961	5.6995	0.0170
A26	1	0.3161	0.1391	5.1640	0.0231
h2	1	0.0938	0.1550	0.3660	0.5452
J1	1	0.4314	0.1186	13.2265	0.0003

Hyväksytty malli:

```
aku06 = vakio + koulu5 + ika3 + kieli + A5 + A26 + J1
```

Mallin tulokset on esitetty taulukossa 4.9.

Mallin Hosmer & Lemeshow -testin arvot kertovat, että malli on merkitsevä:

Chi-Square	DF	Pr > Chi Sq
3.0154	8	0.9334

## b) Kaikki havainnot (n = 3 793)

Täysi malli, josta lähdettiin liikkeelle:

```
aku06 = koulu1 + koulu3 + laani + sukup + ika1 + ika2 + sp_ika1 +
sp_ika2 + A1 + kieli + g89 + g2 + j3
```

**Liitetaulukko 4.** AKU 2006:een osallistuminen (osallistuminen ja vastauskato), logit-mallitus ( $C \sim A_{\text{pitkä, lyhyt}}$ ) täysi malli.

Parameter	DF	Standard		Wald	
		Estimate	Error	Chi-Square	Pr > Chi Sq
Intercept	1	-0.9010	0.3444	6.8423	0.0089
koulu3	1	-0.0968	0.1278	0.5739	0.4487
koulu5	1	-0.4310	0.1437	8.9937	0.0027
Laani	1	-0.1799	0.0956	3.5364	0.0600
sukup	1	-0.1053	0.1435	0.5382	0.4632
ika2	1	-0.5173	0.3922	1.7401	0.1871
ika3	1	-0.4479	0.3357	1.7806	0.1821
sp_ika2	1	0.2297	0.2477	0.8605	0.3536
sp_ika3	1	-0.0201	0.2151	0.0088	0.9254
A1	1	0.4523	0.1201	14.1906	0.0002
kieli	1	-0.4864	0.1658	8.6011	0.0034
G89	1	-0.1329	0.1164	1.3037	0.2535
G2	1	0.1879	0.1161	2.6190	0.1056
J3	1	-0.2043	0.1024	3.9785	0.0461

Hyväksytty malli:

```
AKU06 = vakio + koulu5 + laani + ika3 + A1 + J3 + kieli
```

Mallin tulokset on esitetty taulukossa 4.10.

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > Chi Sq
10.9338	8	0.2055

### Analyysi 3. Aikuiskoulutukseen osallistuminen

Logit-mallituksella selvitettiin, osallistuvatko (AKU 2006:sta) vastaukseen jääneet aikuiskoulutukseen samassa määrin kuin tutkimukseen osallistuneet.

#### a) Selitettävänä muuttujana AKU 2006:een osallistuminen

Täysi malli:

$\text{aku06} = \text{akuOS} + \text{koulu1} + \text{koulu3} + \text{laani} + \text{sukup} + \text{ika1} + \text{ika2} + \text{sp\_ika1} + \text{sp\_ika2} + \text{A1} + \text{kieli} + \text{g89} + \text{g2} + \text{j3}$

**Liitetaulukko 5.** AKU 2006:een osallistuminen ja aikuiskoulutukseen osallistuminen, logit-mallitus ( $C \sim A_{\text{pitkä, lyhyt}}$ ), täysi malli.

Parameter	DF	Estimate	Standard	Wald	Pr > Chi Sq
			Error	Chi-Square	
Vakio	1	1.8395	0.3098	35.2508	<.0001
akuOS	1	-0.0643	0.0507	1.6139	0.2039
koulu1	1	0.8883	0.1074	68.3708	<.0001
koulu3	1	0.5881	0.0839	49.0812	<.0001
Laani	1	0.0248	0.0720	0.1191	0.7300
sukup	1	-0.5887	0.1126	27.3324	<.0001
ika1	1	-0.2107	0.2531	0.6933	0.4050
ika2	1	-0.3789	0.3074	1.5192	0.2177
sp_ika1	1	0.0919	0.1599	0.3307	0.5652
sp_ika2	1	0.1373	0.1952	0.4945	0.4819
A1	1	-1.0000	0.0853	137.3359	<.0001
kieli	1	0.0360	0.1395	0.0664	0.7966
G89	1	-0.2921	0.0889	10.7857	0.0010
G2	1	-0.6884	0.0840	67.1914	<.0001
J3	1	-0.3646	0.0761	22.9638	<.0001

Tarkasteltu paras malli, jossa akuOS pakotettuna mukana, mutta jota ei hyväksytty:

$\text{aku06} = \text{vakio} + \text{akuOS} + \text{koulu5} + \text{laani} + \text{ika3} + \text{kieli} + \text{A1} + \text{J3}$

akuOS-muuttuja tullut tarkastelussa merkitseväksi (esimerkiksi liitetaulukko 6).

**Liitetaulukko 6.** AKU 2006:een osallistuminen ja aikuiskoulutukseen osallistuminen, logit-mallitus ( $C \sim A_{\text{pitkä, lyhyt}}$ ).

Muuttuja	Estimaatti	Pr > ChiSq
vakio	-1,0711	<0,0001
akuOS	-0,1146	<b>0,2474</b>
koulu5	-0,3237	0,0020
laani	-0,1908	0,0440
ika3	-0,4473	<0,0001
A1	0,4685	<0,0001
kieli	-0,4386	0,0045
J3	-0,1968	0,0531

**b) Selitettävänä muuttujana aikuiskoulutukseen osallistuminen.**

Täysi malli, josta lähdettiin liikkeelle:

$\text{akuOS} = \text{aku06} + \text{koul u1} + \text{koul u3} + \text{laani} + \text{sukup} + \text{ika1} + \text{ika2} + \text{sp\_ika1} + \text{sp\_ika2} + \text{A1} + \text{kieli} + \text{g89} + \text{g2} + \text{j3}$

**Liitetaulukko 7.** Aikuiskoulutukseen osallistuminen ja AKU 2006:een osallistuminen (pakotettuna muuttujana), logit-mallitus ( $C \sim A_{\text{pitkä, lyhyt}}$ ).

Parameter	DF	Estimate	Standard	Wald	Pr > Chi Sq
			Error	Chi-Square	
Intercept	1	-1.7589	0.3708	22.4990	<.0001
aku06	1	-0.1241	0.1009	1.5124	0.2188
koul u1	1	0.4078	0.1450	7.9086	0.0049
koul u3	1	0.3172	0.1138	7.7641	0.0053
Laani	1	-0.1810	0.0956	3.5807	0.0585
sukup	1	-0.1113	0.1625	0.4693	0.4933
ika1	1	0.4519	0.3358	1.8112	0.1784
ika2	1	-0.0613	0.4125	0.0221	0.8818
sp_ika1	1	0.0185	0.2151	0.0074	0.9315
sp_ika2	1	0.2477	0.2602	0.9063	0.3411
A1	1	0.4774	0.1218	15.3716	<.0001
kieli	1	-0.4871	0.1660	8.6101	0.0033
G89	1	-0.1252	0.1166	1.1517	0.2832
G2	1	0.2074	0.1172	3.1302	0.0769
J3	1	-0.1957	0.1027	3.6322	0.0567

Tarkasteltu paras malli, jossa aku06 pakotettuna mukana ja jota ei hyväksytty:

$\text{akuOS} = \text{vakio} + \text{aku06} + \text{koul u5} + \text{sukup} + \text{A1} + \text{G89} + \text{G2} + \text{J3}$

Todettiin, ettei akuOS-muuttuja tullut missään tarkastelussa merkittäväksi (katso taulukko 4.21, sivu 65).